Technical Report

UCAM-CL-TR-289
ISSN 1476-2986

Number 289

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

The semantics of noun phrase anaphora

David Alan Howard Elworthy

February 1993

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

This technical report is based on a dissertation submitted
February 1993 by the author for the degree of Doctor of
Philosophy to the University of Cambridge, Darwin College.

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Summary

Anaphora is a linguistic phenomenon in which one expression, called an *anaphor*, gains some or all of its meaning from another, its *antecedent*. In this thesis, I study the semantics of one particular sort of anaphora, where both antecedent and anaphor are noun phrases. Most research in the past has dealt with singular anaphora; I also address plurals.

The two major theories of anaphora are Kamp's Discourse Representation Theory (DRT) and dynamic logics. While they have yielded many valuable insights into the phenomenon, I think it is time to subject them to some critical scrutiny. There are two main critisms. Firstly, the interpretation assigned to the linguistic data is not always consistent with language users' intuitions about it. Secondly, the current theories employ semantic formalisms which rely on either specific representational devices or on unconventional logics. I develop a new theory, TAI (theory of anaphoric information), which attempts to rectify both problems.

The thesis starts with a critical re-examination of the linguistic data, and in particular of the so-called "donkey sentences", which exhibit complex interactions between quantification and anaphora. The following chapter examines DRT and dynamic logics in some detail, considering their successes and failings from both empirical and methodological perspectives.

TAI itself is presented in chapter 4. The theory starts from a conceptual model, which specifies the information needed to interpret anaphors correctly. A logic, L(GQA), is then developed, which derives both truth conditions and the constraints on the anaphoric information from formulae derived from natural language sentences. The logic is static and does not rely on structured representations of the sort found in DRT. The translation procedure from linguistic input to L(GQA) formulae captures a significant part of the empirical weight of the theory, and provides sufficient flexibility to make the required range of readings available.

The last chapter evaluates TAI from a variety of standpoints. The conceptual model is used as a baseline for comparing DRT, dynamic logics and TAI. The relation between semantic properties of TAI and pragmatic aspects of interpreting anaphors is considered. Computational aspects of TAI are also examined: how it relates to Webber's theory of anaphora, and how the logic could be implemented efficiently. Finally, some directions in which research based on TAI could proceed are identified.

Contents

1 Introduction

Anaphora is a linguistic phenomenon in which one expression, called an *anaphor*, gains some or all of its meaning from another, its *antecedent.* Linguists have examined anaphora from a number of different perspectives: at the very least, syntactic, semantic, and pragmatic. In this thesis I am concerned largely with the semantic viewpoint. The central question that is addressed is this: *what is the interpretation of texts containing anaphors?*

There are many different kinds of anaphora; see Hirst (1981) for a comprehensive survey. I restrict my attention to just one, which has variously been called noun phrase anaphora and definite anaphora. Here the antecedent and anaphor are both noun phrases, with the anaphor being either a pronoun or a definite full noun phrase.

During the 1980s, a number of new and important theories of noun phrase anaphora appeared, notably the discourse representation theories of Hans Kamp and Irene Heim, and dynamic logics, which have been most closely associated with Jeroen Groenendijk and Martin Stokhof. While these theories have yielded many valuable insights, I think it is time to subject them to some critical scrutiny. There are two areas where a reassessment is needed. Firstly, the interpretation assigned to the linguistic data is not always consistent with language users' intuitions about it. In particular, they are too rigid in the readings they assign to certain uses of anaphors under quantification. Secondly, the current theories employ semantic formalisms which rely on either specific representational devices or on unconventional logics. While such formalisms may ultimately prove to be necessary, it is not clear that the possibilities of more standard formalisms have yet been exhausted.

Besides the theoretical studies of anaphora, there has also been extensive work in the (overlapping) fields of natural language processing, artificial intelligence and computational linguistics. Neither treatment is sufficient on its own: the theoretical perspective is generally only adequate for highly idealised texts, while the computational work often relies to a greater or lesser degree on principles which lack formal rigour, although there are increasing moves towards examining the computational properties of the logical approaches, and adopting the theoretical standpoints in computational implementations. Most of the work I describe comes from the former tradition, but some reference is made to how practical use might be made of it.

The thesis is organised as follows. Chapter 2 is principally concerned with the data. I start by showing some simple examples of noun phrase anaphora, and gradually introduce more problematic cases, together with some of the ways of classifying and analysing the data which have been proposed. This part of the discussion culminates in the "donkey sentences" of section 2.3. Donkey sentences exhibit complex interactions between quantification and anaphors, and have probably received more attention than any other examples of noun phrase anaphora. Some subsidiary issues concerned with the acceptability of anaphors, of other classes of anaphora, and of the readings of sentences containing plural noun phrases are explored in the remainder of the chapter. In chapter 3, I examine the two theories (or groups of theories) mentioned above in some detail, looking at how they relate to other areas of semantics as well as how they apply specifically to anaphora. My own solution follows in chapter 4. I have tried to take a "grass roots" approach, starting from a general conceptual model and using the linguistic evidence to add detail to it before proceeding to a more formal statement. Finally, I offer an evaluation of my work in chapter 5, looking both at its own merits and how it relates conceptually, formally and empirically to the theories of chapter 3. I believe that the approach I arrive at is computationally viable, both in the sense that it is possible to construct a systematic procedure which takes natural language input to a model-theoretic interpretation, and in the sense that an efficient means of computing with the formalism can be devised. A sketch of the latter also appears in chapter 5.

Many people have helped in the development of my work over the past three years. Ted Briscoe, my supervisor, pointed me in the right direction and made many valuable comments and suggestions. My friends and colleagues in the Computer Laboratory have provided me with feedback and some of the more elusive reference material: Ann Copestake, Antonio Sanfilippo, Dick Crouch, Dave Milward and Karen Sparck-Jones. Alex Lascarides also made a number of useful observations on later drafts of the thesis. Jaap van der Does kindly sent me some copies of papers that were otherwise unobtainable. The TEX macros used for DRSs were provided by Alan Black. I am grateful also to SERC for funding, and to Rank Xerox Cambridge EuroPARC, who provided me with a welcome break from my Ph.D. in the summer of 1991. Finally, Sarah, Rachel and Nicola each played their part.

2 Noun Phrase Anaphora

The goal of a theory of anaphora is to assign appropriate interpretations to texts containing anaphors. In order to know when an interpretation is "appropriate", we first need to arrive at a statement of what such texts mean, based on the intuitions of language users. The first three sections of this chapter address this issue, by examining progressively more complex examples of sentences and discourses involving anaphors. Section 2.1 is concerned with some simple cases, and establishes that anaphora is best analysed semantically, rather than through syntax or pragmatics. The discussion draws on a classification of anaphora following the influential work of Evans (1977, 1980). Most current semantic theories of anaphora acknowledge the value of this classification in understanding the data, although (as the discussion in this chapter and the next will show), some theories have succeeded in applying a uniform interpretation procedure across several classes.

Evans' theory makes an important claim about the interpretation of certain sorts of anaphor: that they refer to a uniquely identifiable object. The point has been contested, and in the second section I examine some of the arguments for and against it.

The first two sections can be seen as preliminary to the main part of this chapter: the discussion of donkey sentences in section 2.3. Donkey sentences involve complex interactions between anaphora and quantification. Even setting aside the issue of how to formalise their interpretation, there is considerable debate about what they mean. If there is a single key point to the chapter as a whole, it is that past work has been too rigid in the meanings assigned to donkey sentences.

A related question is that of when a noun phrase may act as the antecedent for an anaphor. One body of work in this area, arising mostly from theoretical linguistics, is concerned with when the antecedent-anaphor relation is allowed and when it must be blocked. I outline some possible approaches in section 2.4. Computational linguists have been more interested in the issue of anaphor *resolution*; that is, in choosing the most likely antecedent from the ones which are permitted. I do not address this issue here; some references to the relevant literature appear in the section.

Section 2.5 is concerned with some subsidiary issues: anaphora by noun phrases other than pronouns, some classic "difficult" cases, and noun phrase anaphora which requires mechanisms beyond the ones described in the preceding sections. The final topic, addressed in section 2.6, is collective and distributive readings of sentences. While not intrinsically part of the discussion of anaphora, it provides some background for later sections.

I will use the terms *text* and *discourse* more or less interchangeably, to mean a connected sequence of sentences. There is, of course, far more to discourse than this "flat" definition might imply.

2.1 Anaphora: some basic ideas

Consider (1):

(1) Fritz is a cat. He has blue eyes.

If we assume that the antecedent of *he* in the second sentence is *Fritz*, then we could give a meaning to (1) by textually substituting the antecedent for the anaphor, and then interpreting the result:

(1a) Fritz is a cat. Fritz has blue eyes.

It is easy to show that textual copying will not always work. Applied to

(2) A man was standing in the rain. He was unhappy.

textual substitution would yield:

(2a) A man was standing in the rain. A man was unhappy.

which does not mean the same thing: it makes claims about two separate men. Rather than copying the text, we wish to use the identity, or *reference*, of the antecedent to interpret the anaphor. In (2), the the antecedent identifies an individual, which the anaphor then refers to.

Unfortunately, this will not work in all cases either. In (3) (using *s/he* for want of a better gender-neutral pronoun)

(3) Every linguist thinks s/he has the best theory.

s/he does not refer to any single linguist, but to each linguist in turn. This suggests a division of anaphors into two classes. Anaphors of the sort in (2) are called *referential*, because they refer to a specific thing. Those illustrated by (3) are called *bound*, by analogy with the bound variables of logic. That there really is a difference is shown by

(3a) Every linguist thinks s/he knows best. ?S/he is pleased about it.[1]

Here anaphora to *every linguist* is not acceptable outside the scope of *every*, in contrast to (2).

The class of referential anaphors can be further subdivided. In (1), the antecedent brings to mind something that is already known. The same is true in the following examples:

(4) John/The man loves his mother.

(5) John and Bill/The men love their mothers.

In these cases, the antecedent and anaphor can be said to be *co-referential*: they each refer to a given, definite entity. The antecedent could equally introduce a new entity, as it does in (2) or (6).

(6) John owns some sheep and Harry vaccinates them.

Following Evans (1977, 1980), from which (6) comes, anaphors of this sort are termed *E-type*. In E-type anaphora, the antecedent is indefinite, contributing new information, and generally quantified. The new entity introduced by the antecedent in (6) could be described by the (definite) phrase "the sheep John owns". E-type pronouns can be shown not to be bound, by considering a paraphrase (6) would receive if they were:

(6a) Some sheep are such that John owns them and Harry vaccinates them.

---

[1]The notation '?' is used to mark sentences, discourses and lexical items which are generally found unacceptable.

which does not mean the same thing: (6) says that Harry vaccinates all of John's sheep, while (6a) does not.

Anaphora proper may be distinguished from deixis. Anaphors receive their interpretation principally from the *linguistic* context, established by other parts of text they occur in. Deictic expressions draw on a context established non-linguistically. Many of the linguistic forms used to realise anaphora, pronouns especially, may also be used for deixis. An example of a deictic expression is the pronoun in (7), when the utterance is accompanied by, for example, the speaker pointing to someone.

(7) He is beating a donkey.

It is perhaps slightly misleading to talk of anaphors themselves as being divided into classes. What is actually important is the relationship between the anaphor and its antecedent. Compare (8) and (9), which differ only in the antecedent.

(8) John loves his mother.

(9) Every boy loves his mother.

In (8), *his* is co-referential with its antecedent, and in (9) it is bound by it. This implies that we must pay some attention to properties of the antecedent when deciding which class to place an occurrence of an anaphor in. Furthermore, the same antecedent may play different roles with different anaphors. In (10)

(10) Every boy loves his mother. It makes their fathers jealous.

his is a bound anaphor with antecedent *every boy*, while *their* has the same antecedent but is E-type. I will be a little sloppy about this, and refer to classes of anaphor, rather than classes of anaphor-antecedent relationship.

An influential theory of anaphor interpretation is that of Evans (1977, 1980). He considers four classes of pronoun use: deictic, co-referential anaphora, bound anaphora, and E-type anaphora, and notes that there is some overlap between the ways in which different classes are interpreted. In both the co-referential (8) and the bound (9), the verb-phrase can be considered as a predicate of one variable, with the variable substituted for the pronoun. (8) is then interpreted by binding the individual *John* to the variable, and (9) by binding each boy in turn to it. That is, the difference is in what we do with the antecedent, the pronoun being treated uniformly in the two cases. Similarly, there is a connection between co-referential and E-type anaphora. To interpret an E-type anaphor, we identify the whole of what the antecedent refers to and use this as the referent of the pronoun. For co-referential anaphora, the procedure is the same, except that the referent of the antecedent is known prior to the occurrence of the antecedent. Again, the difference lies in the antecedent and not in the anaphor.

Evans contrasts his approach with pragmatic theories of coreference, according to which principles other than the semantic rules of the language are used in explaining the interpretation of anaphors, analogous to using extra-linguistic factors to interpret deictic pronouns. A theory that he considers in some detail is that of Lasnik (1976), who uses a pragmatic strategy to interpret both deictic and co-referential uses of pronouns. In both cases, the pronoun is understood as referring to some entity which has been made salient. The only difference is in how this has come about: linguistically or by extra-linguistic means. Evans has a number of criticisms of Lasnik. Firstly, if co-referential pronouns are interpreted in the same way as deictic expressions, then the similarity between co-referential and bound anaphora is lost, and the fact that the same interpretation procedure can be used in examples (8) and (9) must be considered as accidental. The pragmatic strategy also runs into problems with E-type pronouns, a class of anaphor not considered by Lasnik. According to a pragmatic theory of coreference, *them* in

(11) John owns some donkeys and feeds them at night. (Evans, 1980, p.353)

would have to be taken as referring to some contextually salient group of entities, a likely candidate being the group of donkeys John owns. But if we consider the example of

(12) Every villager owns some donkeys and feeds them at night. (Evans, 1980, p.353)

there is a problem, because *them* refers to a different collection of donkeys for each villager under consideration; there is no single contextually salient group of entities it can refer to. To save Lasnik's theory, we would need one such group for each villager, and to pick the right group as each villager is considered. In contrast, the E-type paraphrase of *them*, "the donkeys the villager owns", captures this directly.

Theories which attempt to reduce most or all uses of pronouns to bound anaphora, most notably those of Geach (1962), also come under heavy criticism from Evans. Example (6a) is one illustration that E-type pronouns cannot be treated as being bound, and a variety of similar examples can be found in Evans (1977, pp.112-122). The most notable of these examples are donkey sentences, the discussion of which is deferred to section 2.3.

A prediction of Evans' theory is that E-type pronouns carry a uniqueness implication. Take example (6), which is predicted to be equivalent to (13):

(6) John owns some sheep and Harry vaccinates them.

(13) John owns some sheep and Harry vaccinates the sheep that John owns.

It appears from (13) that *them* refers to a specific group of sheep, namely the one that John owns, and that there is just one such group of sheep. Similarly in (14a), equivalent to (14b)

(14a) Every villager owns some donkeys and feeds them at night.

(14b) Every villager owns some donkeys and feeds the donkeys he owns at night.

the collection of donkeys is unique relative to each villager. Heim (1982) (reprinted as Heim (1988)) comments in her discussion of Evans' work, that the uniqueness implication should be understood as meaning unique within some domain of quantification, just as quantification in ordinary conversation is normally restricted to some relevant domain. She gives as an example:

(15) A wine glass broke last night. It has been very expensive.

which claims only that some wine glass which broke, unique within a relevant domain, was expensive. Any other wine-glasses that broke were not part of the domain, and any others that were in the relevant domain did not break. This is Heim's reconstruction of how uniqueness is to be understood, and it is not totally clear what Evans' own views are. In Evans (1980, p.343), he quotes

(16) There is a doctor in London and he is Welsh.

as not being acceptable as a means of reporting that there is at least one Welsh doctor in London, on the grounds that the E-type interpretation of *he* carries a implication that there is a unique doctor in London. Heim, on the other hand, is willing to admit that *a doctor* could refer to a unique individual, given a suitable context, and that (16) would then be acceptable; for example, if it were preceded by another speaker saying

(16a) This boy is sick, but he speaks only Welsh and we can't understand his complaints.

Unfortunately, this line of argument brings us back to a point where we rely on pragmatic factors (or at least, unexplained notions of context) to define the domain of quantification for (16). If this is so, the uniqueness implication need not be a consequence of the E-type pronoun: it is built into the domain that is being considered as relevant. This criticism is, in part, behind the rejection of Evans' analysis of anaphora which was led by Heim and others. In section 2.3 I will come back to this issue. However, I first turn to another discussion of uniqueness, which will also have a bearing on the data of section 2.3.

Digression: E-type or D-type?

Evans (1977) explicitly rejects the view that E-type pronouns go proxy for descriptions, in the sense that they may be interpreted by generating a suitable description from the antecedent at some level of linguistic representation and substituting the description for the pronoun. His preferred view is that E-type pronouns have their referents fixed by a description: that is, a description derived from the antecedent defines some collection of individuals, and the anaphor also refers to that collection, using the same mechanism as co-referential anaphors. This view is discussed by Neale (1990, pp.184-189), who uses examples of pronouns in modal and temporal contexts to show that there is a need for pronouns which do go proxy for descriptions. For example in

(17) Boston has a mayor. He used to be a Democrat. (Neale, 1990, p.187)

there are two possible interpretations of the second sentence. On Evans' non-proxy approach, the referent of *he* is fixed by the definite description *the mayor of Boston*, interpreted in the temporal context of the first sentence, and so (17) says that the person who is now mayor of Boston used to be a Democrat. But an equally valid interpretation is that the person who used to be mayor of Boston was a Democrat, but that the current mayor may or may not be. In this case, *he* can be taken as going proxy for the definite description, in that substituting it into the second sentence gives the correct interpretation. Neale therefore suggests that there is another class of pronoun for the latter case, which he calls D-type. After some consideration, he concludes that E-type pronouns can be reduced to a special case of D-type ones. Outside modal and temporal contexts, the E-type/D-type distinction is not of great importance, and for the discussion here I will continue to refer to E-type pronouns alone, with some vagueness about which mechanism of interpretation is the correct one.

2.2 Uniqueness and definiteness

Kadmon (1990) presents an extensive discussion of what uniqueness means and what triggers it. The issue itself goes back at least far as Russell (1905), in which the idea is put forward that definite descriptions, including anaphors, carry a uniqueness implication, and that this is what distinguishes them from indefinites. The property of uniqueness applies to the referent of the NP in question. For singulars, this object is an individual, as in

(18) The king visited me.

For plurals NPs, the object is a plural collection described by the NP, which is only unique if it is the largest such collection. Thus, (19) says that *all* the kings visited me:

(19) The kings visited me.

Similarly, in (6),

(6) John owns some sheep and Harry vaccinates them.

them refers to the maximal collection of sheep owned by John.

Kadmon reports Russell's view on uniqueness as having been challenged by Strawson (1950) and Searle (1969). According to Strawson, in

(20) The table is covered with books.

there clearly is a unique table which is the one referred to, but the sentence may still be used felicitously if there is more than one table known to the speaker and hearer. Examples of this sort therefore appear to go against the claim that the definite NP must have a unique referent. Nevertheless, Kadmon and many others[2] believe that uniqueness can be salvaged if additional factors are allowed in determining the referent of a definite NP, such as the contextual limitation of the domain seen in (16). Kadmon calls this approach *realistic uniqueness*. She summarises her views as follows:

> I find that careful examination of the data supports the generality of uniqueness and confirms a realistic version of Evans' predictions. I would like to argue for this generalization: definite NPs which are not syntactically (c-command) bound must stand for a set/individual uniquely identified by some property known to the language user. (Kadmon, 1990, pp.278-279).

Furthermore, she claims that discourse anaphora not only display the uniqueness effect, but act as the source of it. That is, uniqueness effects appear only when there is an anaphor. Not everyone agrees: Kadmon notes that van Eijck, Webber and Reinhart, amongst others, attribute the uniqueness effect to the nature of the indefinite antecedent NP.

In support of her view, she looks first at some simple examples, such as

(21) Leif has a chair. It is in the kitchen.

Uttered in isolation, the first sentence of (21) does not imply either that Leif has only one chair, or that the speaker has a specific one of Leif's chairs in mind. It is only on interpreting the second sentence that the uniqueness implication arises, and unless there is a way of singling out one of Leif's chairs, many speakers will find (21) unfelicitous. The same effect occurs with plural indefinites, as in

(22) Leif has four chairs. They are in the kitchen.

[2]See Kadmon (1990, p.276) for references.

The uniqueness effect is sometimes obscured by other factors. Kadmon examines the situation where Leif has ten chairs, all of which are in the kitchen. A uniqueness account claims that neither (21) nor (22) is acceptable in this situation, but some speakers nevertheless do find them felicitous. Kadmon's explanation is that such cases are acceptable only if there is nothing which would distinguish one chair, or set of four chairs, from another. The speaker still does assert that the pronoun has a unique referent, but that the referent could be any of the equivalent possibilities. In contrast, when (say) seven of the ten chairs are in the kitchen, and the remainder elsewhere, their location acts as a distinguishing property, and neither (21) nor (22) is acceptable. Heim's wine glass example, (15), can perhaps be explained by similar principles.

One further refinement to uniqueness is needed. In (23), from Sells (1985),

(23) Every chess set comes with a spare pawn. It is taped to the top of the box.

we do not want *it* to imply that there is just one pawn, but that there is a unique spare pawn for each chess set. Kadmon therefore introduces the idea of *relative* uniqueness: that under quantification, uniqueness applies to each of the things – chess sets, in (23) – that is quantified over.

The conclusion from Kadmon's analysis is that definite NPs can be interpreted as referring to a unique entity, provided the definition of uniqueness is stated carefully (although Kadmon's own formalisation is not of this sort). Consequently, an E-type analysis of certain pronoun occurrences is possible, in which the pronoun is interpreted as a definite description and therefore accompanied by a uniqueness implication. However, we have so far looked at the interaction of quantification and anaphora only in a superficial way. In the next section I examine sentences which may suggest that neither an E-type analysis nor any of the other classes suggested is appropriate.

2.3 Donkey sentences

Donkey sentences provide critical data for theories of anaphora, in that they exhibit a complex interaction between antecedents and anaphors under quantification. Heim (1982) defines them as

> ...sentences that contain an indefinite NP which is inside an if-clause or relative clause, and a pronoun which is outside that if-clause or relative clause, but is related anaphorically to the indefinite NP. (p.44)

The indefinite may also appear in a PP modifier. Modern interest in donkey sentences originates with Geach (1962). The two canonical examples are:

(24) Every farmer who owns a donkey beats it.

(25) If a farmer owns a donkey, he beats it.

I will refer to sentences of similar form to (24) and (25) as *quantified* and *conditional* donkey sentences, respectively.

The pronouns in (24) and (25) cannot be co-referential, because there is no single thing that they refer to. Nor can they be treated as bound, since the pronouns lie outside the scope that would conventionally be given to their antecedents. For example, assuming a compositional translation into first-order predicate calculus, with pronouns translated into the variable bound by their antecedents, (24) would yield:

(24a) $\forall x[(farmer(x) \wedge \exists y[donkey(y) \wedge own(x,y)]) \rightarrow beat(x,y)]$

where the y in $beat(x,y)$ is free. Similarly, for (25), the translation is

(25a) $(\exists x \exists y[farmer(x) \wedge donkey(y) \wedge own(x,y)]) \rightarrow beat(x,y)$

in which both variables are free in $beat(x,y)$. This factor distinguishes donkey sentences from examples such as (26) with conventional translation (26a), analysing the possessive by means of a predicate *of*.

(26) Every boy loves his mother.

(26a) $\forall x[boy(x) \rightarrow \exists y[mother(y) \wedge of(x,y) \wedge love(x,y)]]$

In this case, the variable x for *his* is bound by the universal quantifier.

The only remaining class of pronominal anaphor is E-type. To see whether it is applicable, we need to look at what donkey sentences mean and whether the interpretation of the pronoun is consistent with the uniqueness implication. There has been considerable discussion on this issue. Many linguists and philosophers of language have taken both (24) and (25) to mean the following:

> **Universal reading**
> Every farmer who owns one or more donkeys beats every donkey which he owns, i.e.
> $\forall x \forall y[(farmer(x) \wedge donkey(y) \wedge own(x,y)) \rightarrow beat(x,y)]$

The universal reading (sometimes called the strong reading) does not follow from the E-type analysis, on a simple statement of it at least, since *it* does not refer to a unique donkey per farmer. If donkey sentences are to be assigned a universal reading, some other semantic mechanism must be used for the anaphor. However, not everyone agrees that the universal reading is the correct one. In the following sections, I will examine some alternative readings and the complex array of arguments and counterarguments that have been put forward.

One caveat: the data about donkey sentences must, in my opinion, be taken with a pinch of salt. Most of the discussion has centred around a small number of examples, which have been so extensively examined that it is hard to arrive at any very definite intuitions about them. A scan of the LOB corpus, which contains approximately one million words of natural British English text, revealed only a single instance of a "real" quantified donkey sentence:

(27) At midday a pelting shower soaked the ground: the thirty men moved off across the field to their dinner, and as they went, *every foot, treading on a hoed-up weed, planted it again in the receiving earth.* (G19 76-79)

The intended interpretation of this sentence is not easy to determine.

2.3.1 Quantified donkey sentences

2.3.1.1 Universal and other readings

Consider the following examples, attributed to Barbara Partee:

(28) Every man who has a daughter thinks she is the most beautiful girl in the world.

(29) Every man who has a son wills him all his money.

Cooper (1979) says of (28):

> It does not seem to me, at least, that [(28)] commits any father of more than one daughter to the contradictory belief that each of his daughters is the most beautiful girl in the world.

Similar remarks may be made for (29): it does not entail that a father of more than one son leaves his entire fortune to each of them. Another example is (30) (assuming monogamy), which contrasts with (31), where a universal reading is possible.

(30) Every man who loves a woman marries her.

(31) Every man who loves a woman likes her.

Cooper uses (28) as an argument against the universal reading. Before looking at his alternative analysis, we may note that the universal reading can be forced in some contexts. For example:

(32) Every man who has a son wills him all his money. I wonder how Joe's two sons, Bill and Fred, will react when they find out he's double-crossed them.

Cooper proposes that the meaning of (24) and similar sentences may be represented as

(33) $$\forall u[(\mathit{farmer}(u) \wedge \exists v[\mathit{donkey}(v) \wedge \mathit{own}(u,v)]) \rightarrow \exists x[\forall y[[S(u)](y) \equiv y = x] \wedge \mathit{beat}(u,x)]]$$

(in fact, this slightly simplifies Cooper's formula.) S is a relation between farmers and the objects which the pronoun may represent. Paraphrasing (33): for all farmers u who own a donkey, there is exactly one object x of which $S(u,x)$ holds, and which u beats. By choosing different values for S, different readings can be obtained. If it is taken as being the logical translation of the definite description *the donkey he owns*, the reading is that there is just one donkey owned by each farmer, for *the* to be felicitous. The sentence says nothing about farmers who own more than one donkey. I will call this the unique antecedent reading. It is the one predicted by Evans' analysis:

Unique antecedent reading
Every farmer who owns a donkey beats the donkey he owns.
$$\forall x[(\mathit{farmer}(x) \wedge \exists! y[\mathit{donkey}(y) \wedge \mathit{own}(x,y)]) \rightarrow \exists! y[\mathit{donkey}(y) \wedge \mathit{own}(x,y) \wedge \mathit{beat}(x,y)]]$$
(where $\exists! y[P(y)]$ is equivalent to $\exists y[P(y) \wedge \forall z[P(z) \rightarrow z = y]]$

Alternatively, we can choose S such that a farmer may own more than one donkey, exactly one of which is beaten by him:

Unique anaphor reading
Every farmer who owns a donkey beats exactly one donkey he owns.
$$\forall x[(\mathit{farmer}(x) \wedge \exists y[\mathit{donkey}(y) \wedge \mathit{own}(x,y)]) \rightarrow \exists! y[\mathit{donkey}(y) \wedge \mathit{own}(x,y) \wedge \mathit{beat}(x,y)]]$$

Either choice of S will solve the problem of (28). Cooper's preferred reading is the unique antecedent one, on the grounds that the factors that would enable us to single out one donkey in the unique anaphor case are unclear. He also comments that he has no clear proposal for the pragmatic factors that would lead to selection of an appropriate function for S.

Heim (1982) re-examines Cooper's analysis, starting from a reconstruction of what she takes the reasoning behind it to be. She suggests it runs as follows: if the universal reading is adopted, then (28) will always be false in a world where there are men who have more than one daughter. However, sentences like (28) do not immediately strike one as being automatically false, and so there must be an implicature (presupposition) which admits the possibility of such sentences being true. A suitable implicature is that the quantification is restricted to fathers of exactly one daughter. The uniqueness condition in (33) is a consequence of the implicature, and the choice of S is independent of this implicature. Furthermore, the implicature follows from the presence of the pronoun: contrast

(34) Every father has a daughter

where there is no need to make the implicature.

Heim raises two criticisms, on the assumption that this is the reasoning behind Cooper's theory. First, she questions the apparent assumption that implicatures, or the combination of an implicature and an entailment are relevant in picking the analysis, but logical entailments alone are not. There are examples where a similar limitation to the domain of quantification does appear to follow from an entailment. In

(35) John thinks that nobody is as smart as him.

the domain over which *nobody* quantifies must exclude John, since there would otherwise be a contradiction, that John thinks John is not as smart as John. In this case, the domain limitation arises from the entailed contradiction, rather than an implicature arising from the pronoun. Heim's argument on this point is a little unclear, but what I think she is saying is this. (35) shows that in some circumstances, there is a need to avoid contradiction by limiting the domain of quantification. So, the reading initially assigned to (28) could be the universal one, and when it is seen that this entails a contradiction, the domain limitation is applied, resulting in Cooper's reading. What Cooper had proposed was that the pronoun led to an implicature which similarly limited the domain, but this was done *grammatically*, instead of as a pragmatic response to an entailed contradiction. On Heim's account, the universal reading is primary and the unique one is arrived at by entailment; on Cooper's the unique reading is primary and arrived at by implicature.

Some support for Heim's claim can be added as follows. If we try treating the domain restrictions as presuppositions – one form of implicature – then it fails a diagnostic described by Levinson (1983). The test, which is not infallible, consists of considering both the sentence and its negation. For the domain restriction (or any other inference) to be considered a presupposition, it must be entailed by both. In (28), the presupposition we are interested in is something like

(36) Only men with exactly one daughter are under consideration.

The sentential negation of (28) is

(28a) It is not the case that every man who has a daughter thinks she is the most beautiful girl in the world. = Some man who has a daughter does not think she is the most beautiful girl in the world.

I do not think that this presupposes (36), and so the uniqueness effect can, as Heim says, be considered an entailment rather than an implicature.

Heim has a second argument, based on her "sage plant" sentence. Unlike the previous argument, which was directed against unique antecedent readings, it also applies to unique anaphor ones. We are asked to imagine a situation in which sage plants are always sold in flats of nine. Then the following sentence is certainly true:

(37) Everybody who bought a sage plant here bought eight others along with it.

For Cooper's analysis to hold, it is necessary to assume that *it* singles out one of the nine sage plants for a given customer, an assumption which seems implausible. Heim comments:

> Imagine somebody uttered [(37)], and you asked them to make explicit which one of every buyer's sage plants they meant. It would not feel like an appropriate question to ask. (p.89)

Heim's conclusion is to reject the Cooper analysis, and to develop a theory which yields only the universal reading. An example which is in some ways related to the sage plant sentence is (38), from Neale (1990).

(38) Every farmer who owns more than one donkey beats it.

As with the sage plant examples, there is more than one donkey per farmer, and nothing singles out one as the referent of *it*.

Although Heim's criticisms of uniqueness seem sound, the conclusion that the universal reading must be adopted as primary does not necessarily follow. One answer to it will appear when we look at how Kadmon's uniqueness theory applies to donkey sentences. Another possibility is that there is an ambiguity in the interpretation of the indefinite article *a*. Sometimes it is interpreted so that *it* in donkey sentences picks up all things it might refer to, hence leading to a universal reading; on other occasions, it either only allows *it* to pick up a single thing (unique anaphor), or alternatively carries a genuine uniqueness condition (unique antecedent). Evidence from determiners other than *every* goes some way to supporting this possibility, as discussed below.

Cooper is not the only person to have presented data against the universal reading. Schubert and Pelletier (1989) quote

(39) Every man who owns a donkey will ride it to town tomorrow.

which they claim to be true if every male donkey-owner will ride as least one of his donkeys to town tomorrow. Equally, they have an example against unique antecedent readings:

(40) Everyone who has a donkey must donate its services for one day during the festival.

They comment:

> Does anyone seriously think that, if [(40)] were ordered by the local government, wealthy farmers with two or more donkeys could plead that they were exempt on that ground alone. And against the universal reading, surely these wealthy farmers are not required by this order to donate the services of *all* their donkeys? (p. 200).

Schubert and Pelletier propose yet another possible reading, which they call the Indefinite Lazy Reading. For the canonical quantified donkey sentence it is:

> **Indefinite Lazy reading**
> Every farmer who owns one or more donkeys beats one or more of the donkeys he owns.
> $$\forall x[(farmer(x) \land \exists y[donkey(y) \land own(x,y)]) \rightarrow \exists y[donkey(y) \land own(x,y) \land beat(x,y)]]$$

They consider this reading to be the one that "comes closest to capturing the intuitions behind these examples". The universal reading can be explained as no more than an extreme case of the indefinite lazy reading. Schubert and Pelletier's paper is concerned principally with generics, and they note that something similar happens in generic (or habitual or gnomic) sentences:

(41) If I find a quarter, I'll give it to you.

can have a meaning where I give you every quarter I find, but it need not be as strong as this.

All of Schubert and Pelletier's examples involve modal or temporal auxiliaries: *will*, *must*, etc. Without the auxiliaries, the indefinite lazy reading is far less apparent. For example,

(42) Every man who owns a donkey rides it to town.

contrasting with (39), can quite naturally have a universal reading, perhaps with each man riding different donkeys on different occasions but eventually riding all of them. One example for which the indefinite lazy reading is available without such auxiliaries is (43), which may be contrasted with (30) and (31):

(43) Every man who loves a woman kisses her.

which I find can be read as saying that every man who loves one or more women kisses some of the women he loves.

In spite of this, it is not totally clear to me that the indefinite lazy reading does exist. One way of explaining it is that the sentence is actually being given a unique anaphor reading, so that in (42) each farmer can own more than one donkey, but that one specific one is being ridden to town. The appearance of there being more than one donkey that is ridden to town arises because there is some indeterminacy in which donkey is chosen as the referent of the anaphor. This might help to explain why the indefinite lazy reading is clearer with modal and temporal auxiliaries: they provide an index for identifying which is the donkey in question. A test to determine whether the reading does exist might be to continue with a sentence containing an anaphor to the whole collection of donkeys:

(44) Every farmer who owns a donkey will ride it to town tomorrow. They are taking them to be vaccinated.

The second sentence implies that *all* the donkeys are to be vaccinated, and not just one per farmer. However, it may be that the analysis of (44) is a matter for the semantics of anaphora in temporal contexts, rather than the interaction of anaphors and quantifiers.

Many of the theories of anaphora of the last decade have adopted the universal reading, often without detailed consideration of the alternatives. Besides Heim (1982), they include Groenendijk and Stokhof (1987), Barwise (1987), Muskens (1991), and the Discourse Representation Theory of Kamp (1981). The more important of these theories are discussed in the following chapter. Rooth (1987) largely adopts the same position, while noting some counterexamples. For example, if there is one farmer who beats nine out of the ten donkeys he owns, and all other farmers beat all of their donkeys, then

(24) Every farmer who owns a donkey beats it.

is judged true by some informants and false by others, whereas

(45) Every donkey which is owned by a farmer is beaten by him.

is always judged false, even though the universal reading as formulated above predicts that the two sentences have the same truth conditions.

Rooth's data suggests that it may be necessary to take the topic of the sentence into account. The context established by the rest of the discourse can also make a difference, as in the following example from Chierchia (1991, p.54):

(46a) In Ithaca, every farmer who owns a donkey beats it.

(46b) The farmers of Ithaca are stressed out. They are constantly arguing and often even beat each other. To put an end to it, they go to the local psychologist who recommends that rather than beating each other, every farmer who owns a donkey should beat it. They follow her advice, and things improve.

The indefinite lazy reading appears to be best here: it is not of interest how many donkeys are beaten by each farmer.

2.3.1.2 Other determiners

When quantified donkey sentences with determiners other than *every* are considered, more problems arise. Assume, for the moment, that the universal reading is applicable to other determiners. The universal reading for sentences with *every* can be restated as

Universal reading for *every* (restatement)
$\forall(x, y)[\mathit{farmer}(x) \wedge \mathit{donkey}(y) \wedge \mathit{own}(x, y), \mathit{beat}(x, y)]$

meaning that all pairs of x and y satisfying the first condition in the brackets also satisfy the second condition. Taking

(47) Most farmers who own a donkey beat it.

it is tempting to state the universal reading as:

Universal reading for *most* (incorrect version)
$MOST(x, y)[\mathit{farmer}(x) \wedge \mathit{donkey}(y) \wedge \mathit{own}(x, y), \mathit{beat}(x, y)]$

interpreted as (something like) at least half the pairs of x and y satisfying the first condition also satisfy the second one. However, as Rooth (1987) points out, if there are a hundred farmers, one with a thousand donkeys, all of which he beats, and ninety-nine with one donkey, which none of them beat, then the sentence is intuitively false, but is predicted to be true by the above formula, since there are 999 pairs of farmers with beaten donkeys and 99 pairs of farmers with non-beaten donkeys. This has been dubbed the proportion problem. However, it is easy enough to arrive at a formulation which does work:

Universal reading for *most* (correct version)

$MOST(x)[farmer(x) \wedge \exists y[donkey(y) \wedge own(x,y)],$
$\qquad \forall y[donkey(y) \wedge own(x,y) \wedge beat(x,y)]$

i.e. most farmers who own one or more donkeys beat all the donkeys they own.

Assuming the right formulation is made, there is the more important question of whether the universal reading is the correct one. Heim (1982) makes the comment that for (47), she finds her intuitions vacillate between a universal and a unique anaphor reading. I find much the same thing: the sentence is making a claim primarily about farmers, and whether one donkey is to be considered or many is of lesser significance. Similar remarks can be made about many other determiners: there simply isn't a clear meaning. Other than *every*, the only determiner for which there has been widespread acceptance of the universal reading is *no*. A frequently quoted example is (48), from Rooth (1987):

(48) No parent with a son still in high school has ever lent him the car on a weeknight.

It is claimed that (48) disallows any reading except the one where no parent lends his car to any of his sons. The universality of *no* can also be defeated, as it is when (48) is followed by

(49) However, some parents with two sons will do so, because one son keeps an eye on the other.

It is striking that *every* and *no* tend to encourage the universal reading more than other determiners, and it may be because they make claims about all objects in the domain.

2.3.1.3 Kadmon's uniqueness theory and donkey sentences

Kadmon (1990) has applied her uniqueness theory to donkey sentences, using an analysis which she describes as being close in spirit to Cooper's. As noted in the previous section, her central view is that definite NPs, including discourse anaphora, are systematically accompanied by uniqueness implications, and that the uniqueness is a consequence of the definiteness of the NP. The sentence

(50) Most women who own a dog talk to it.

is interpreted by Kadmon as quantifying over women who own a unique dog, which she claims is in accordance with most speakers' intuitions. In contrast,

(51) Most women who own a dog are happy.

carries no uniqueness implication, since there is no definite NP which would require it. Uniqueness is to be understood as meaning that there is some factor that would allow one dog to be singled out for each woman:

> For example, if we are talking about animals that are kept indoors, [(50)] might be about women who have exactly one dog that they keep indoors. The point is that [(50)] is about women who each have a dog which is unique in some way. (Kadmon, 1990, p.307)

Where context forces consideration of women who own one or more dogs, Kadmon reports that speakers are often unable to decide on a truth value for (50). The canonical donkey sentence (24) is given a similar reading. Kadmon comments:

> My intuition is that I have no idea what [(24)] says about what men with two or more (non-unique) donkeys do with their donkeys. This judgement has been confirmed by several informants. I think that [(24)] cannot be used felicitously if the domain of quantification contains men with non-unique donkeys. (Kadmon, 1990, p.314)

In my classification, Kadmon's analysis is a unique antecedent one.

Kadmon deals with Heim's arguments against uniqueness as follows. She first notes that her intuitions about both (24) and (47) are insecure, and that her uncertainty arises directly from the restriction of the domain to those individuals meeting the uniqueness requirement. The paraphrase of the universal reading, *every farmer who owns one or more donkeys...* specifically directs attention to a different domain, and difficulties in interpreting donkey sentences arise from the conflict between the two conceptions of the domain. When the universal reading does appear to be the best one, Kadmon attributes this to the "influence" of (25):

(25) If a farmer owns a donkey, he beats it.

which, as discussed below, can be given a universal reading quite naturally. Kadmon claims that (25) encourages thinking about *instances* of farmers beating donkeys they own, and that some parallel with (24) leads us to think of it in the same way. There is no such parallel for (47), which explains why it is harder to obtain a universal reading. As an explanation of the sage plant example (37), one of Heim's counter-examples to uniqueness, Kadmon suggests that there are obscuring factors, of the "Leif's chairs" sort (example (21) of the previous section). Specifically, if there is no way of distinguishing the sage plants, then the one we pick out to make unique is arbitrary, the sentence having the same truth conditions whichever choice is made.

While I agree with Kadmon's analysis of the data as suggesting that unique readings are available, I am unconvinced by either of Kadmon's arguments for her specific standpoint on how uniqueness comes about. There is little syntactic similarity between (24) and (25), so the required parallel must be found at some other level. Saying it is in the semantics renders the whole argument circular. Furthermore, there is no reason why a parallel between donkey sentences with *most* and conditionals of the form

(52) Mostly, if a farmer owns a donkey, he beats it.

should not be constructed, even though the latter is (on some readings) consistent with a universal interpretation of

(47) Most farmers who own a donkey beat it.

The uniformity argument about sage plant sentences is also weak. As Kadmon describes it, the speaker always assumes that the sage plant is unique per buyer, but the choice of which sage plant is singled out is left undetermined, since it does not affect the truth conditions (Kadmon, 1990, p.317). I am not sure this is true. If there is a subsequent inter-sentential reference, as in

(53) Most women who own a dog talk to it. They never talk back.

then *they* refers to all the dogs which were actually used in determining the truth conditions of the first sentence: all dogs owned by the women on a universal reading, one per woman on a unique anaphor one, and so on. Examples of this sort also seem to undermine Kadmon's position that the uniqueness arises from the anaphor. Consider

(54) Most women who own a dog are happy. They talk to them.

According to Kadmon, there was no need for a uniqueness implication in the first sentence. But I find the two sentences taken together have exactly the same range of readings as the first sentence of (53): that most women talk to one dog they own, or that they talk to all of them. A solution to both these problems is to drop the idea that uniqueness comes from the anaphor, and locate it in the anaphoric properties of the indefinite determiner, making it (as already suggested) ambiguous between passing on all possible objects to anaphors, and passing on one specific one.

All of the examples considered so far have used indefinite NPs that are singular: *a donkey*, *a daughter*, and so on. Similar properties can be observed with plural indefinites, for example,

(55) Every farmer who owns four donkeys beats them.

Kadmon reports that informants who disagree with her about uniqueness in (24) are often much more willing to accept uniqueness with numeral determiners. This may be attributable to an ambiguity in the determiner between *exactly four* and *at least four*. In the former case, the universal and unique anaphor accounts coincide. If the *at least* is made explicit, I find the universal reading is more natural, but both are possible. Non-numeral plural indefinites, as in

(56) Every farmer who owns some/several/many donkeys beats them.

favour a universal reading, to me at least. Kadmon does not consider these data.

2.3.1.4 The relationship between the readings

One way of understanding how the readings relate to one another is as follows. We first need the following sets:

A = farmers who own exactly one donkey.

B = farmers who own more than one donkey.

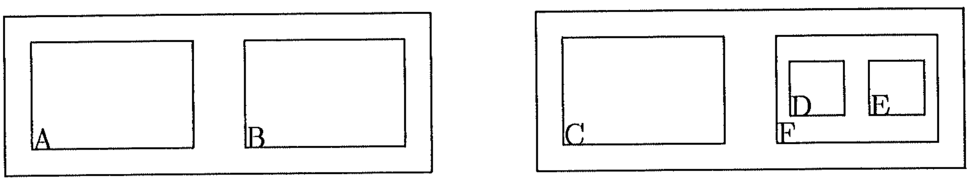
C = farmers who own exactly one donkey and beat it.

D = farmers who own more than one donkey and beat exactly one of them.

E = farmers who own more than one donkey and beat all of them.

F = farmers who own more than one donkey and beat one or more of them.

A and B are disjoint, as are C and F, and D and E, and D and E are subsets of F, i.e. the relations between the sets are:



own   own and beat

The truth conditions of the readings are equivalent to:

Universal: $A = C$ and $B = E$

Unique antecedent: $A = C$

Unique anaphor: $A = C$ and $B = D$

Indefinite lazy: $A = C$ and $B = F$

On the basis of these conditions, we can treat the unique antecedent reading as being the logically weakest one, in the sense that its truth does not preclude the truth of the other readings. The next weakest is the indefinite lazy reading. Neither the universal reading nor the unique anaphor reading is weaker than the other. A different perspective is to say that the indefinite lazy reading is the "base" reading, and that the others are produced from it by applying some sort of strengthening principle. To obtain the unique antecedent reading, the principle is to limit the domain of quantification. For the unique anaphor and universal readings, extra conditions must be placed on the interpretation of the anaphor: that it refers to one donkey per farmer, or to all donkeys per farmer. The constraints could be driven by reasoning of the sort already mentioned, such as the need to avoid a contradiction in Cooper's (28), or by uniformity across the objects involved, as in Kadmon's analysis of the sage plant example (37). I will return to this issue in section 2.3.3.

2.3.2 Conditional donkey sentences

Conditional donkey sentences are exemplified by

(25) If a farmer owns a donkey, he beats it.

Many semantic theories have treated conditionals as material implication, i.e. to say that *If A, then B* holds except in the case that A is true and B false. The arguments against such an approach originate with Lewis (1975), who points out that adverbs of quantification may be used to restrict the relation between the antecedent and the consequent of the conditional.[3] Thus, to quote two examples from Heim (1982),

(57) If a table has lasted for fifty years, then it will always last for another fifty years.

(58) If a table has lasted for fifty years, then it will sometimes last for another fifty years.

the conditional can be treated as a quantification over cases (in a sense to be discussed) where there is a table that has lasted for fifty years and cases where a table that has done so will last for a further fifty. (57) asserts that all cases of the former sort are, or can be extended to, cases of the latter sort; (58) that some cases of the former sort are of the latter sort. Sentences without an explicit adverb of quantification may be treated as having an implicit adverb roughly equivalent to *always* or *generally*. Some supporting evidence for treating conditionals in this way comes from parallels with sentences where the quantification is more explicit. Thus, (59) parallels (25), and (60) parallels (61) and (62).

[3] There is an unfortunate overloading of the term "antecedent" between the predecessor of an anaphor and the left hand side of the conditional. It will generally be clear which is meant in the course of this section.

(59) Whenever a farmer owns a donkey, he beats it.

(60) If a farmer owns a donkey, he usually beats it.

(61) When a farmer owns a donkey, he usually beats it.

(62) A farmer who owns a donkey usually beats it.

For the purposes of this section, I will concentrate on sentences of the form *If A, B.*

Lewis' way of obtaining the universal reading of (25) is to treat the antecedent and consequent of the conditional as being open sentences, in which the variables do not have explicit quantifiers. Cases are equated with assignments, i.e. mappings from the free variables to individuals. The interpretation rules can then be expressed in terms of quantification over assignments. For *always*, a suitable rule is (as Heim (1982, p.125) quotes it):

> "always(ϕ, ψ)" is true iff every assignment to the free variables in ϕ which makes ϕ true also makes ψ true.

What is important is that the quantification is over a sequence of individuals, one for each NP in the antecedent; in (25), for example, over pairs of individuals. Chierchia (1991), following Kadmon, calls this the *symmetric* reading, and suggests that there are also *asymmetric* readings, one for each indefinite in the antecedent. He uses the example

(63) If Mary lends a book to a student, he returns it with good comments.

According to one possible reading, the topic of (63) is Mary's students, and the reading is

(63a) If Mary lends A BOOK to a student, he returns it with good comments (while if she lends them a tape, they don't know how to work with it).

Chierchia describes (63a) as the indirect object asymmetric reading, since it the indirect object over which the quantification takes place. There is also a direct object asymmetric reading:

(63b) If Mary lends a book to A STUDENT, it gets returned with good comments (while if Mary lends one to a colleague, it gets heavily criticised).

The difference between the readings can be expressed in terms of what the implicit adverb of quantification applies to: the books for (63a) and the students for (63b).

The availability of symmetric and asymmetric readings has been widely noted when the quantification is made explicit, as in

(64) If a farmer owns a donkey, he usually beats it.

A symmetric interpretation rule for *usually* is:

> "usually(ϕ, ψ)" is true iff at least half the assignments to the free variables in ϕ which make ϕ true also make ψ true.

This reading runs into the proportion problem that was observed for the first version of the universal reading for (47) given above.

(47) Most farmers who own a donkey beat it.

A more natural reading of (64) is the subject asymmetric one, according to which a claim is made about most farmers and not most farmer-donkey pairs. By contrast, (25) favours the symmetric reading, although it is possible (with some difficulty) to give it an asymmetric reading.

An alternative to treating bare conditionals as having an implicit adverb of quantification is to postulate an implicit necessity operator. Kadmon (1990) makes essentially this distinction, using the terms *multi-case* conditional and *one-case* conditional respectively. Multi-case conditionals quantify over multiple cases or instances of the antecedent sentence, as in (65):

(65) If a man walks in and he sits down, Sally is pleased.

where the quantification is over instances of a man walking in and sitting down. (65) says that in all cases of a man walking in and sitting down, then necessarily, Sally is pleased. One-case conditionals are exemplified by (66) and (67):

(66) If there is a doctor in London and he is Welsh, then we are all set.

(67) If it is true a man walked in and that he decided to stay, then Sally will be pleased.

In (66), the consequent is true if the state described by the antecedent is true. Only one way of making the antecedent true need be considered. Similarly, the consequent of (67) is claimed to be true whenever the events of the antecedent have already taken place. Many conditionals may be interpreted either way.

Kadmon's claim is that one-case conditionals show a uniqueness effect: indefinite NPs appearing in them are to be taken to refer to one specific individual. Thus Kadmon says that (66)

> ...is to be understood as being about the possibility that there is exactly one doctor in London and that doctor is Welsh. (p.298)

As with the quantified donkey sentences, we may again have to have some additional information to single out a unique doctor. Multi-case conditionals do not carry uniqueness implications. Thus, Kadmon's

(68) If a semanticist hears of a good job, she applies for it.

does not imply that either the semanticist or the job are unique, nor that there is a unique job per semanticist. Nevertheless, it is possible to use a uniqueness analysis for multi-case conditionals, and this analysis also provides a means of obtaining the asymmetric readings as well as the symmetric ones, as we shall see in a moment.

Kadmon's uniqueness analysis is not the first such attempt. Cooper (1979), as discussed above, and Evans (1980) constructed theories which are based on presuppositions that indefinites refer to unique individuals. In her discussion of such theories, Heim (1982) quotes

(69) If a man is in Athens, he is not in Rhodes.[4]

[4]Heim attributes this example to the Stoic philosopher Chrysippos.

which cannot adequately be understood as meaning that there is exactly one man in Athens, even when relativised to a possible world and/or a time, and which may therefore appear to be evidence against uniqueness. Heim (1982) suggests salvaging Cooper's and Evans' approaches by making the uniqueness be relative to the "cases" over which the conditionals quantify, as suggested above. There is some difficulty in doing so; all the suggestions she puts forward for what might constitute a case run into problems. The contribution of Kadmon (1990) is to propose a notion of case that does seem to work, and a later work of Heim (1990) suggests how it could be formalised.

Heim's reconstruction of Kadmon's theory is this: multi-case conditionals quantify over situations, which are to be taken as partial worlds. Situations may be extended by adding extra information, provided no inconsistency results in the extended situation. Adverbs of quantification are interpreted as relations between the set of situations in which the antecedent is true, and the set of extended situations for which the consequent is also true. For (25), situations for the antecedent contain one farmer and one donkey owned by the farmer, and the farmer and donkey distinguish one situation from another. The consequent adds the additional information to the situation that the farmer must beat the donkey. To obtain the universal reading, we require that the first set be a subset of the second. Asymmetric readings are obtained by formulating the situations differently. For the subject asymmetric case, the sets of situations are distinguished only on the farmer. All that we require is that there is some donkey in the situation which he owns and (for the consequent situations) beats. The relation expressed by the adverb of quantification is the same as before. A similar approach yields the object asymmetric reading. Generalisation to examples with more than two indefinites is straightforward.

If we adopt this approach, Kadmon's uniqueness claim holds for all conditional donkey sentences, whether one-case or multi-case, symmetric or asymmetric. However, a significant question remains: how are we to decide when a one-case reading is appropriate, and when a multi-case, and further which of the possible multi-case readings to take? One possibility is to leave the whole issue vague, and admit all of the possible interpretations, but this is unsatisfactory, since some of the examples quoted above do seem to favour or disfavour certain readings. Heim (1990) reports a suggestion made by Baüerle and Egli (1985), according to which the situations are to be distinguished by those antecedents for which anaphors appear in the consequent. Thus, for

(70) If a farmer owns a donkey, he usually deducts it from his taxes.

the symmetric reading is appropriate because anaphors to both antecedents appear in the consequent, whereas for

(71) If a drummer lives in an apartment complex, it is usually half empty.

the object asymmetric reading should be taken. Both Heim and Kadmon have questioned the empirical adequacy of this: other factors, such as topic-focus structure appear to make a difference as well. For example, Heim quotes the following, where capitalisation represents stress:

(72) Do you think there are vacancies in this apartment complex?
Well, I heard that Fulano lives here, and if a DRUMMER lives in an apartment complex, it is usually half empty.

(73) Drummers mostly live in crowded dormitories. But if a drummer lives in an APARTMENT COMPLEX, it is usually half empty.

In Heim's view (72) makes a claim about apartment complexes and (73) about drummers. The former is consistent with Baüerle and Egli's suggestion, the latter not. Furthermore, their approach actually leads to the reading of (64) which suffers the proportion problem. Heim suggests that Baüerle and Egli's proposal may in part arise from presuppositions of the sort discussed for quantified donkey sentences. That is, when a singular pronoun appears in the consequent, there is a presupposition that the antecedent for the pronoun determines a unique individual. This explanation does advance the discussion a little, but, as Heim notes, it still does not help with the focus and proportion problems.

There is one final form of conditional donkey sentence which has caused some problems with a uniqueness approach, exemplified by (74) and (75):

(74) If a bishop meets somebody he blesses him.

(75) If a man shares an apartment with another man, he shares the housework with him.

Suppose that the *somebody* of (74) was another bishop: then neither *he* nor *him* can be taken as referring to *the unique bishop who met another bishop.* Kadmon's approach does not have the same problem, since the uniqueness is not derived purely from the sentence containing the antecedent. To interpret (75), we would pick one man and nominate him as unique, and then pick another unique man who lives with the first man.

Heim raises a problem in a similar "indistinguishable antecedent" case. Consider

(76) If a man has the same name as another man, he usually avoids addressing him by name.

Suppose, to take Heim's example, that there are 100 men called John, all of whom avoid addressing all others by name, and 200 men not called John, each of whom shares his name with exactly one other man, and none of whom avoids addressing any other by name. Then the subject asymmetric reading predicts the sentence is false: the men called John, who are non-"name-addressers" are in the minority. On a symmetric reading, the sentence should be true: there are 9900 ordered pairs of men called John satisfying the consequent, and 200 ordered pairs of men with the same name (not John), who falsify it. Heim notes that she is unsure what the relevant judgements are, and I find myself in the same position. The most obvious reading to me is that the sentence tells us nothing about the actions of arbitrary pairs of men in general. What is does say is that on *occasions* when two men meet, then if they have the same name, more often than not neither of them uses it. The quantification by *usually* is not over pairs of men at all, but over instances of men meeting, which may include the same two men meeting more than once. It should be possible to integrate this into the situation based account given above, by distinguishing situations on the basis of events rather than individuals.

2.3.3 Summary and conclusions

Quantified donkey sentences

To recap: there are four possible readings of donkey sentences: universal, unique antecedent, unique anaphor and indefinite lazy. For each of the readings, there are some sentences which appear to make that reading the most natural one, and some which seem to forbid it. Take four examples quoted earlier:

(28) Every man who has a daughter thinks she is the most beautiful girl in the world.

(37) Everybody who bought a sage plant here bought eight others along with it.

(40) Everyone who has a donkey must donate its services for one day during the festival.

(47) Most farmers who own a donkey beat it.

The suggested readings are:

1. Universal: favoured by (37), opposed by (28).
2. Unique antecedent: favoured by (47), opposed by (37).
3. Unique anaphor: favoured by (28), opposed by (37).
4. Indefinite lazy: favoured by (40), opposed by (28).

Furthermore, many factors seem to contribute to how readings are favoured or opposed: the predicates involved, the determiner of the subject NP, and the context of occurrence of the sentence.

The approach that has been taken by most theories is to settle on one reading and ignore the others. I think this is misguided: it is simply not consistent with the data. There are two alternatives. Note first that the critical factor in obtaining the different readings is the interpretation of the indefinite article in *a donkey* and of the pronoun *it*, or the corresponding items in other examples. One solution is to make these items ambiguous, so that all of the possible readings are generated. In some cases, such as (28), one or more of the readings will be impossible in that they lead to a contradiction. The second possibility is to take one reading as the standard one, used by default, and to allow a different reading to be used instead, on the basis of other factors of some sort. If a contradiction results, then the reading is abandoned and a different one used instead. This solution can be seen as being a constrained variant of the first in that we do not allow all of the ambiguities unless there is some factor which triggers them, and that the alternatives are placed in order, with the "best" one selected unless it is blocked. The second account is more in keeping with a definition of ambiguity expressed by Barbara Partee (quoted in Kálmán and Szabó (1990, p.262)):

> An expression that has several meanings qualifies as ambiguous if and only if the speaker is not entitled to be uncertain about the particular reading of the expression (s)he has in mind when uttering it. The expression has an underspecified meaning otherwise.

In the case of donkey sentences, one meaning does usually stand out as the intended one. By underspecified, Kálmán and Szabó mean that

> different readings stem from different possibilities of interpreting its floating conditions (p.261)

which in this case means the interpretation of the indefinites and anaphors.

It is difficult to give a definitive answer to the question of which reading is the standard one, and what factors trigger other readings. Lexical items certainly do seem to have an effect. For example, as Heim noted, *every* leads more readily to the universal reading than *most* does. Emphasis and stress also make a difference. If stress is placed on *a*, as in

(77) Every farmer who own A donkey beats it.

I find the unique antecedent reading strongest, as it is in the contextually modified (78).

(78) Every farmer who owns a donkey beats it, but ones who own two do not.

Similarly, the context in (79) favours the universal reading (to me), in contrast to Schubert and Pelletier's prediction.

(79) The council needs as many donkeys as it can get this year. Everyone who has a donkey must donate its services for one day during the festival.

A final example is

(80) A farmer who owns a donkey beats it.

which is most naturally read as being about one farmer and his one donkey, but which can be modified into universal or unique antecedent readings:

(81) Generally, a farmer who owns a donkey beats it.

(82) A farmer (whom I have in mind) beats a donkey whenever he owns it.

Conditional donkey sentences

For conditional donkey sentences, readings may be classified as one-case or multi-case, and within the latter, as symmetric, or as asymmetric on each of the antecedents. Some sentences favour one reading, others are neutral between some or all of the readings. As with quantified donkey sentences, an ambiguity account could be adopted, or we could generate the alternative readings with an ordering on them. A variety of factors affect which reading is the most prominent one, including (at least) explicit adverbs, topic, and context of occurrence. Further complications arise from examples like (76), which seem to favour quantification over events rather than individuals. One solution to this is to use a mixed approach, where there is ambiguity between the event based and non-event based interpretations, and then ambiguity or underspecification within the latter.

2.4 Accessibility, acceptability and agreement

An account of anaphora must specify not only how anaphors are to be interpreted, but which NPs are suitable and possible as antecedents for a given occurrence of an anaphor. The question of anaphor resolution, or choosing the best antecedent from the available possibilities, has been extensively discussed in the computational literature; see, for example, Grosz (1986), Sidner (1986), Webber (1979), Hobbs (1986) and Carter (1987). I am concerned with a different issue: determining when an antecedent-anaphor relation must be prohibited for reasons which are (largely) independent of the detailed properties of the referents. When the antecedent-anaphor relation is possible, I will talk of the anaphor *accessing* the antecedent and of the antecedent being *accessible* to the anaphor. To illustrate, in (83), the second occurrence of *it* is found unacceptable by most speakers.

(83) Every farmer who owns a donkey beats it. ?It brays in distress.

The antecedent *a donkey* is therefore said to be inaccessible to the second occurrence of *it*.

There are three main strategies for determining when the antecedent-anaphor relation is allowed:

Syntactic: use lexical properties of the antecedent and the anaphor, together with configurational information, such as their relative positions in the parse tree.

Structural: use configurational properties of the semantic representation, independently of the interpretation of that representation.[5]

Semantic: use information that can be derived purely in the interpretation process. Barwise (1987) calls this a notion of a semantically admissible assignment of antecedent-anaphor relations.

The structural and syntactic accounts may not be as distinct as they first appear. If a purely compositional procedure is available for translating from parsed sentences to their representations, structural configurations could be traced back to all of the possible parse trees that result in them, although it may be inconvenient to express the constraints in these terms. A semantic account of acceptability differs in being able to draw on extra information, such as predicate extensions in a given model.

Most current theories of anaphora use structural constraints. Two notable examples are Kamp's Discourse Representation Theory (DRT), and Groenendijk and Stokhof's Dynamic Predicate Logic (DPL), discussed in detail in chapter 3. The essential idea in both cases is that the antecedent introduces some sort of discourse referent or variable into the semantic representation, and the anaphor is translated into the same discourse referent. Some constructs, including negation, conditionals and universal quantifiers, have the effect of making any discourse referents introduced within them unavailable outside the construct. In (83), the discourse referent arising from *a donkey* is accessible within the scope of *every*, i.e. the first sentence, but not outside it. Examples of scope limitation by negation and conditionals are:

(84) Joe doesn't own a donkey. ?It lives in a field.

(85) If a farmer owns a pedigree donkey, he is rich. ?It lives on caviar.

In (84), no discourse referents within the negation may be accessed from outside it. Similarly, in (85), the conditional blocks further reference to the discourse referents for both farmer and donkey.

The structural constraints of DRT and DPL are usually expressed in terms of syntax of the semantic representation they use, in a manner which is independent of the interpretation of the representations. Other kinds of structural account are possible; Chierchia and Rooth (1984) give an example for DRT, as follows. The semantic representations used in DRT (discourse representation structures – DRSs) are given a truth conditional interpretation by relating them to a model through embedding functions, which map each discourse referent to an individual of the model. The embedding functions are partial, the discourse referents they are defined on being the same as the ones which are accessible

[5] I am using representation to mean the formal notation into which the sentences are translated. It should not be confused with representation in the sense of something that might be psychologically real. The issue of representation is addressed in section 3.2.1.

from the DRS being embedded. An attempt to access any other discourse referent will block, because the embedding function is undefined on that value. Chierchia and Rooth thus manage to do away with the need to define accessibility directly in terms of DRS syntax, by saying that when interpretation blocks for this reason, the sentence is unacceptable. However, this is still a structural notion of acceptability, in the sense that the domain over which an embedding function is defined derives from the structure of the semantic representation, and not from the particular objects which embedding functions map discourse referents to.

The situation becomes more complex with plural data:

(86) Every farmer owns a donkey. They bray/?it brays in distress.

The structural approaches described above will make *it* unacceptable. To allow *they*, there must be some means of forming a new discourse referent representing the collection of donkeys owned by the farmers, which is accessible outside the first sentence, and requiring that only plural anaphors can access that discourse referent. Furthermore, the new discourse referent must not be accessible within the first sentence, as (87) shows:

(87) Most farmers who own a donkey beat it/?them.

If a plural NP is used instead of *a donkey*, there is a plural discourse referent which is accessible within the first sentence, but with a different interpretation to the one outside it:

(88) Every farmer who owns two donkeys beats them/?it. They bray in distress.

Problems of this sort can be solved within a structural account, as the discussion of plural DRT in section 3.1.4.2 will show, at the cost of introducing some complexity into the theory.

However, there are some cases where structural theories may run into difficulties. The acceptability of an anaphor can be contingent on the model with respect to which the sentence is interpreted. Kamp and Reyle (1990) consider the following:

(89) Every director gave a present to a child from the orphanage.

(89a) One of them opened it/its present straight away.

(89b) One of them opened ?them/?their presents straight away.

(89c) Two of them opened them/their presents straight away.

Kamp and Reyle's claim is that (89a) is acceptable, just as (89c) is, but that (89b) must be rejected. However, I think this results from our preconceptions about the model, that there are fewer directors than there are children. (90) is structurally the same as (89), and yet (90a) now seems less good and (90b) rather better:

(90) Every child gave a present to a teacher from the school.

(90a) One of them opened ?it/?its present straight away.

(90b) One of them opened them/their presents straight away.

The preconception this time is that there are more children than teachers. At least one teacher must therefore have received more than one present, and that teacher can be the subject of (90b). That is, a structural definition of acceptability is not sufficient in these cases, because the acceptability depends on what the model happens to contain.

A related phenomenon appears with numerically specified anaphors. Suppose that on hearing

(91) Some people are standing at the bar.

it can be identified that three people are meant. Then (91a) is acceptable, but (91b) is not.

(91a) The three of them are drinking beer.

(91b) The four of them are drinking beer.

Again, this is contingent on the situation. Taken independently of a particular model, there is no way of predicting the acceptability.

The fact that the acceptability can be contingent on a model might suggest looking to a semantic account. On this approach, a collection of individuals is associated with each antecedent (its referent) and anaphors express constraints on that information. An anaphor is unacceptable if its constraints are not satisfied by the information available. For example, *it* might be expected to require that the antecedent information consists of a single thing, and similarly *they* requires more than one thing: that is, there is *semantic number agreement.* Semantic number agreement will explain (89)–(90). This does not show that a structural account is wrong, but only that it is not sufficient in all cases. A purely semantic account will work is sufficient for most of the examples seen so far, repeated here:

(83) Every farmer who owns a donkey beats it. ?It brays in distress.

(84) Joe doesn't own a donkey. ?It lives in a field.

(85) If a farmer owns a pedigree donkey, he is rich. ?It lives on caviar.

(86) Every farmer owns a donkey. They bray/?it brays in distress.

In each of (83), (84) and (86), the unacceptability of *it* arises because its referent is not singular. There is more than one donkey (assuming more than one farmer) in (83) and (86), and none in (84). The same applies to (85): the conditional does not guarantee the existence of either one donkey or many.

There are two cases where some care is needed to make semantic number agreement work. Firstly, in

(92) Every child got one present and two sweets.
One of them unwrapped the present/?the presents and ate the sweets/?the sweet.

the agreement on *the present(s)* and *the sweet(s)* must apply not to the whole of the antecedent, but to just that part of it relevant to the child indicated by *one of them.* This may be termed *agreement under restriction*, in that the information from the antecedent is restricted by the identity of the child before the agreement test is made. Note also how the agreement changes when there are two children, and hence two presents:

(93) Every child got one present and two sweets.
Two of them unwrapped ?the present/the presents.

In this example, the collection of presents relevant to the two children is plural. Another construct that may trigger agreement under restriction is the floating quantifier *each*, as in

(94) Every farmer owns a donkey. They each beat it.

where *it* is acceptable because *each* restricts the VP to each farmer in turn.

The other difficult case for semantic number agreement is exemplified by donkey sentences. Taking two earlier examples,

(87) Most farmers who own a donkey beat it/?them.

(88) Every farmer who owns two donkeys beats them/?it.

In these cases, we may not want semantic number agreement to apply at all, even under restriction. Supposing that (87) is assigned a universal reading, then *it* is acceptable even though it may refer to more than one donkey for each farmer. Similarly, *them* is always unacceptable, whether there is one donkey or several. What (87) and (88) suggest is that in some cases, purely syntactic agreement is required, and that in these cases, the semantic number agreement is suppressed. Something similar has been proposed for donkey sentences by Neale (1990). He suggests that antecedents and anaphors must agree in syntactic number, but that pronouns are ambiguous between a number-marked and a numberless version. A weakness of Neale's account is that it introduces an extra ambiguity on pronouns, and it is possible to do better. One possibility is to look for a syntactic relation between the antecedent and anaphor which triggers syntactic agreement over semantic. As further data, consider

(95) Most shepherds who know every farmer who owns a donkey beat ?it.

(96) A banker gives every farmer who leases a farm its freehold.

(97) Every farmer who owns a donkey knows a shepherd who beats it.

(98) A banker gives its freehold to every farmer who leases a farm.

(99) Every farmer who beats it hates his donkey/?his donkeys.

(assuming the cataphora in the latter two examples is allowed.) The critical example is (95). In all of the other cases, the following relation holds between the antecedent NP (NPa) and the anaphoric NP (NPp): the NP which most immediately dominates NPa is a sister node to a node dominating NPp. In (95), this is not so, and hence it is marked as unacceptable. The NP node which most immediately dominates the antecedent is *every farmer who owns a donkey*, and the anaphor is not dominated by a sister of this node. Syntactic agreement must also apply in cases traditionally classed as bound anaphora:

(100) Every cat washes itself.

(101) Every cat washes its kittens.

The main difficulty in stating the condition in these terms is that quantifier scope must be taken into account. For example, (95) becomes acceptable if the scope of *a donkey* is widened.

If we are looking to non-semantic principles to support a semantic account, it is tempting to ask whether a purely syntactic account would succeed instead. Syntactic number in itself isn't enough for (83) and some of the other examples above. In these cases, we need to do something along the following lines: trace a path through the parse tree from antecedent to anaphor. If the path passes through certain sorts of nodes, modify the syntactic agreement features. For example, passing through a sentence node for which the subject NP has *every* as its determiner causes singular agreement markers to be changed to plural. Some NPs, such as singular proper nouns, must not be modified. The rule must be applied after scope disambiguation, so that if *a donkey* is given wide scope in (83), then *it* is acceptable in the next sentence. The partitive in (89a) appears to allow either singular or plural pronouns in the object NP, which forces a further modification of syntactic agreement. Also, semantic agreement must make some contribution, for the contingent acceptability in (89)–(90). Another point to note is that the agreement condition must make use of a different feature from the one used in syntactic determiner-nominal agreement. In particular, *every* must be marked as singular for determiner-nominal agreement and as plural for antecedent-anaphor agreement. The conclusion is that not only does syntactic agreement need some rather *ad-hoc* principles, but additional semantic principles are sometimes needed.

There are two minor factors that complicate the data. Firstly, *they* and its case variants may be used as singular. Hirst (1981) calls this the *singular epicene* use. It is typically used where the gender is unknown or deliberately left unmarked. An example is:

(102) Every farmer likes their donkey.

with *their* acting as a bound anaphor. (102) is acceptable only if a singular epicene use of *their* is assumed.

Secondly, there are differences in judgement over reference to NPs quantified with *every*, as in (103).

(103) Every farmer owns a donkey. They beat them.

Intuitions about the acceptability of the second sentence vary between language users. Presumably this arises because *every farmer* is syntactically singular but describes a plural collection of individuals, whereas *they* is both syntactically and semantically plural. People who judge (103) unacceptable may be reacting to the syntactic number mismatch; people who accept it are willing to drop this given the semantic number agreement. The distance between the antecedent and the anaphor may make a difference to acceptability: (104) is less acceptable than (105), to some speakers at least.

(104) Every farmer owns a donkey. They are cruel.

(105) Every farmer owns a donkey and lives on a farm with a barn. They are cruel.

However, the data on the phenomenon are unclear, let alone how the syntactic and structural/semantic principles interact.

2.4.1 Subordination

A further complication arises from subordination[6], illustrated by (106)–(108).

(106) Every rice-grower in Korea owns a wooden cart. He uses it when he harvests the crop. (Sells, 1985)

(107) Every rice-grower in Korea owns a wooden cart. Usually, it is a rickety old thing. (Sells, 1985)

(108) If John bought a book, he'll be home reading it by now. It'll be a murder mystery. (Roberts, 1987)

To interpret these examples, we effectively allow the scope of the quantifier in the first sentence to be extended over the second. So, for example, (106) can be paraphrased as *every rice grower in Korea owns a wooden cart and uses it when he harvests the crop.* Subordination is subject to some restrictions, as (83) and the following examples (from Sells and Roberts) show.

(106a) Every rice-grower in Korea owns a wooden cart. ?He also owns a large drying-shed.

(107a) Every rice-grower in Korea owns a wooden cart. ?It is a rickety old thing.

(107b) Every rice-grower in Korea owns a wooden cart. ?Usually, he is poor.

(108b) If John bought a book, he'll be home reading it by now. ?It's a murder mystery.

Sells and Roberts discuss some factors which license subordination. According to Sells, the crucial notion is *discourse continuity*, which is realised in a variety of ways. (106) illustrates temporal subordination: there is an implicit temporal quantifier such as "in each such case". (107) is an example of centering, i.e. continuity of what the discourse is about. In (108), there is modal subordination, indicated by a continuity of mood: both sentences are about situations in which John buys a book. In each case, the precise rules that license subordinating the second sentence to the first are difficult to determine. Both Sells and Roberts attempt to do so, but there are some examples where the general principle is not obvious. (109) does not seem to permit subordination, but (110) does, despite the structural parallels between the sentences. Opinions about the acceptability of (110) vary with speaker. It is perhaps easier to get a reading from it when it is taken as, say, a rule of a game rather than a description of an actual situation.

(109) Every man walks in the park. ?He whistles.

(110) Every player chooses a pawn. He places it on square one.

Subordination is only possible with some determiners. For example (111) does not seem to be acceptable.

(111) Five rice-growers in Korea own a wooden cart. ?He uses it when he harvests the crop.

[6]Often referred to as modal subordination in the literature, following the work of Roberts (1987). As the examples show, the subordination may be of other sorts also.

The same is true if any other plural subject NP is used, which may indicate that subordination to the subject is only possible with a NP that is syntactically singular but semantically plural: in English NPs with *every* as determiner. Subordination on other antecedents in the sentence is possible:

(112) Most farmers own a donkey. It is usually old and mangy.

Subordination also occurs in conditional examples, such as

(113) If a client arrives, you treat him politely. You offer him a cup of coffee.

Again, there appears to be some sort of discourse continuity in the second sentence.

An adequate formulation of the licensing rules is likely to require a better notion of discourse structure and rhetorical relations than is available at present. Both structural and semantic theories can handle accessibility in subordination examples. On a structural theory, the representation of the second sentence is somehow inserted into that of the first sentence, in the case of (106)–(108) within the scope of *every*, and in (113) within the representation of the consequent. This is what Roberts (1987, 1989) does, in a DRT framework. On a semantic theory, the collection of individuals from the antecedent over which subordination occurs is taken, and then the subordinated sentence is interpreted with respect to each individual in the collection, applying semantic number agreement under a restriction to this individual.

2.4.2 Acceptability

A concept that has been left vague in the discussion so far is what it means for an anaphor to be unacceptable. On a syntactic or structural account, an unacceptable sentence is rejected with no attempt to interpret it. For the semantic account, we have to decide whether truth and acceptability form part of the same interpretation process, or whether there are two distinct parts to the semantics, one yielding truth conditions and the other acceptability conditions. At first sight, there does seem to be a difference. On hearing either of

(114) Some men were walking in the park. ?He was whistling.

(115) One man was walking in the park. ?They were whistling.

it is easy to decide that the anaphor in the second sentence is not acceptable if its antecedent is the subject of the first one, without knowing what the world (or the described situation) is actually like. In contrast, to decide whether a sentence is true or not does appear to require such information: taken in isolation from a model of the world *Some men were walking in the park* cannot be judged as either true or false. It thus looks as if the conditions for acceptability and the conditions for truth are different sorts of things, with the implication that they should be formalised differently.

However, it is not as simple as this. There are cases where truth is independent of a model, and where acceptability is dependent on one. Given

(116) Joe, who is a man, is not a man.

the truth or falsehood of the sentence could be evaluated by constructing its truth conditions, and then testing them against a model. But it is not necessary to do so: the properties of the logical connectives involved tell us that the sentence can never be true regardless of what model we have. In a similar vein, there are cases such as

(117) Joe, who is alive, is dead.

which would normally be recognised as false because we know that (conventionally) for something to be alive it cannot be dead, regardless of any actual enumeration of all the things that are alive and are dead. It is possible that the same thing is happening when an anaphor is judged as unacceptable, so that when (114) is rejected, it is for the same reasons that (116) or (117) are judged as being always false: it involves a contradiction between the information contributed by *some men* and that contributed by *he*, regardless of the identity of the individuals involved. What this means for a semantic account of acceptability is that the truth and acceptability parts of the formalisation need not necessarily be kept distinct. Instead, acceptability can be recognised, if it needs to be, by a similar mechanism which spots the oddness of (116): that is, by a logical contradiction.

2.4.3 Summary

Three general approaches to accessibility, that is, constraints on the anaphor-antecedent relation, have been proposed, using syntactic, structural and semantic relations. The syntactic account makes use of agreement features on lexical items and phrases. It runs into difficulties with determiners like *every*, which have to be marked with separate number features for determiner-nominal agreement and for antecedent-anaphor agreement, and in cases where one determiner modifies the number feature of another. The structural account has received a detailed investigation in DRT and dynamic logics, and bases agreement on configurational relations in the semantic representation or on semantic properties corresponding to the interpretation of these relations. In the case of plural data, the structural account of DRT has required some additional syntactic and semantic constraints, the latter to deal with acceptability that is contingent on a model. The third possibility is semantic number agreement, which makes the agreement test depend on properties of the referent of the antecedent and anaphor. It puts more of a burden on the semantics, requiring a semantic characterisation of the anaphoric information and a notion of agreement under restriction. An additional syntactic or structural constraint is needed for the special case of "donkey" anaphora. So far as I know, there has been no attempt to develop an account based principally on semantic number agreement. The theory presented in chapter 4 employs the approach outlined here.

2.5 Other kinds of anaphora

All of the anaphors discussed so far have had the following characteristics:

1. They have been pronouns, i.e. have little or no semantic content of their own.
2. Their antecedents have been single NPs.
3. They have been textually preceded by their antecedents.
4. They have obtained their meaning from the referent of the antecedent.

In this section, I survey some examples of anaphora where these characteristics do not hold.

2.5.1 Full NP anaphors

Pronouns, with their low semantic content, are ideal for studying general properties of anaphora, since extraneous factors are kept to a minimum. When full (non-pronominal) NP anaphors are used, some more complex phenomena emerge. An example of a full NP anaphor is *the man* in

(118) A man was standing in the rain. The man was miserable.

Full NPs may be anaphoric only when they are definite; (119) is not anaphoric:

(119) A man was standing in the rain. A man was miserable.

Since indefinites are generally used to introduce new objects into the discourse, and definites for adding information about given ones, we assume that the second *a man* in (119) does not refer to the individual already mentioned. In (118), the anaphor and the antecedent provide the same information about the individual. The anaphoric NP can also add information. In

(120) A dog lived in a kennel. The mongrel used to howl at night.

the mongrel tells us more about the dog. A related example, from Hirst (1981, p.13) is

(121) Ross used his Bankcard so much, the poor guy had to declare bankruptcy.

where the epithet *the poor guy* indicates the speaker's attitude towards the referent. A common use of full NP anaphors is to bring information back into focus when there is intervening material:

(122) I know a man with two talented monkeys. They can sing, dance and canvass for the Conservative Party. He also owns four elephants, but he's never managed to get them to do any tricks. They/The monkeys live in his house.

They most naturally refers to the elephants. The definite NP allows the previously mentioned monkeys to be referred to without having to repeat the full description of them.

Where the antecedent refers to more than one individual, a full NP anaphor can refer to part of the collection, as in

(123) Some children were playing. The boys were teasing the girls.

I will term such cases *partial* and the ones where the whole of the antecedent is taken *complete*. Partial full NP anaphors cannot add any new information about the selected individuals (though the comment part of the sentence, i.e. the VP, may do so). The contrast can be seen in (124) and (125).

(124) John was playing. The tall boy was happy.

(125) Some children were playing. The tall boys were happy.

In (124), *the tall boy* is complete, and tells us additional information about John. In (125), *the tall boys* either provides the additional information that all the children were tall boys (if complete), or it selects just those individuals who were boys and were tall (if partial). It cannot refer to some subset of the children and also add the new information that the members of the subset are tall, for the obvious reason that there would be no way of determining which subset was meant.

In (123) and the partial reading of (125), the manner in which we selected which subset of the antecedent to use was straightforward: we just took the individuals from the antecedent which satisfied the full NP. Kamp (1990, pp.121-122) gives a sequence of examples which show the means by which the referent of the anaphor is inferred may be quite complex, possibly involving some knowledge of the internal structure of the antecedent. To identify the referent of *the man* in the examples that follow, the correct member of the antecedent must be deduced using either world knowledge, as in (126) and (127), or the gender of the possessive pronoun, as in (128) and (129).

(126) A sculptor and his son walked into a pub. The older man was wearing a grey overcoat.

(127) A priest and a young woman walked into a pub. The man was wearing a grey overcoat.

(128) A sculptor and his spouse walked into a pub. The man was wearing a grey overcoat.

(129) A sculptor and her spouse walked into a pub. The man was wearing a grey overcoat.

Clark and Clark (1977, p.97) introduce the idea of bridging assumptions, implicatures which are required to infer a suitable antecedent for a full NP.

(130) Patience walked into a room. The chandeliers burned brightly.

To resolve the full NP anaphor *the chandeliers*, a hearer adds the implicature

(130a) The room referred to by *a room* had chandeliers in it.

Similar examples (not from Clark and Clark) are:

(131) Bill was buried yesterday. The gravedigger spilt his beer on the coffin.

(132) Joanna's new car was standing in the driveway. The windows were made of black glass.

In each case, some sort of prototypical information is being used: that burying involves a gravedigger in (131) and that cars have windows in (132). Bridging assumptions are also needed with some occurrences of pronominal anaphors, for example

(133) The ham sandwich at table three wants a coke. He's getting impatient.

as uttered in a restaurant.

Full NP anaphora is still subject to the accessibility constraints described in the previous section. Neither of the following is acceptable on narrow scope readings of the indefinite (assuming the second sentence is not subordinate to the first).

(134) Every woman walked into a room. ?The room was brightly lit.

(135) Every woman walked into a room. ?The chandelier burned brightly.

The theories described in this and the next chapter make little attempt to account for any of these examples. Complete full NP anaphors are straightforward to deal with. An account that also covers partial full NPs and the kinds of inference seen in (131) and (132) is rather harder, and it seems likely that a general theory will need to draw on some kind of world knowledge, as well as using the semantics and accessibility mechanisms involved in simpler kinds of anaphora.

2.5.2 Complex antecedents

Anaphors can draw on a combination of several NPs for their antecedent, as in the following, both of which are taken from Kamp (1990).

(136) John took Mary to Acapulco. They had a lousy time.

(137) Last month John took Mary to Acapulco. Fred and Suzie were already there. The next morning they set off on their sailing trip.

In both cases, the anaphor refers to individuals drawn from several NPs, even though there is no syntactic conjunction. Note that which NPs are combined can be affected by context. The natural reading of (137), where *they* refers to all four individuals, can be changed by preceding it with

(138) John is having an affair with Fred's wife.

The operations used to form the antecedent are not unconstrained. A classic example by Partee shows that a "set subtraction" operation is not allowed.

(139) I lost ten marbles and found all but one of them. It is probably under the sofa.

(140) I lost ten marbles and found only nine of them. ?It is probably under the sofa.

(139), with an explicit antecedent *one of them* is acceptable. (140) is not permitted, the putative antecedent being the marbles that are not described by the VP of the first sentence. A suitable bridging assumption would make it possible. The unacceptability implies that language users find such an assumption unreasonable.

2.5.3 Cataphora

In all of the examples so far, the antecedent textually precedes the anaphor. Another sort of reference is possible, termed *cataphora*, in which the pronoun comes first.[7] Cataphora tends to be far more restricted than backward anaphora, usually occurring intrasententially, as in (141) (from van Deemter (1990)).

(141) Ever since her childhood, Dorit has been extremely lazy.

Cataphora across sentences is possible, as in:

(142) First he lost his wallet. Then his car got stolen. Fred was having a bad day.

[7]Note that the term "anaphora" is sometimes used to cover both backward reference, i.e. anaphora proper, and forward reference, i.e. cataphora.

Van Deemter comments that the grammaticality of (142) and the naturalness of the reading are debatable (although it is common as a literary device). Cataphora does not seem to require any further classes of anaphor over the ones presented already. The major challenge is to define conditions on when cataphora may be used, and possibly on when it is preferable to backward anaphora. An example of the need for the latter comes from (143), which is less felicitous than (141).

(143) Ever since Dorit's childhood, she has been extremely lazy.

(141) is also in contrast to (144),

(144) ?Ever since her childhood, a girl has been extremely lazy.

which is unacceptable, although it is easier to get a reading if *a girl* is taken as meaning "a specific girl". This suggests that there is a need for some sort of definiteness constraint. A similar point is made by Barwise (1987).

A comprehensive theory of cataphora is likely to need to appeal to notions of topic structure, and perhaps to some measure of processing complexity. Topic structure might explain (141) compared to (143). Dorit is most likely to be the topic, and so one might expect the subject of the main sentence to be *Dorit* and not an expression referring to her. Processing complexity may help explain the poor naturalness that some people find with (142). The first two sentences cannot be fully interpreted until the third one is reached and the cataphors resolved, and hence some sort of "partially evaluated" structure may need to be retained. Both of these issues lie beyond the scope of the present work.

Mutual anaphora

One further sentence form is of interest: Bach-Peters sentences, of which (145) is an example.

(145) A boy who deserved it got a prize he wanted.

Deemter calls such examples *mutual anaphora*, since each noun phrase contains an anaphor to the other, one backward and the other cataphoric. Sentence (145) was originally put forward as evidence that "textual copying" would not work as a mechanism for all cases of anaphora. Copying the antecedents in place of the pronouns yields

(145a) A boy who deserved a prize he wanted got a prize a boy who deserved it wanted.

and trying to get rid of *he* and *it* in the result makes the expanded form keep on growing. A bound anaphor approach will not work either, because *it* does not fall within the scope of its antecedent. An Evans analysis will work, although it is rather more convoluted than the examples we have already seen. For (6),

(6) John owns some sheep and Harry vaccinates them.

it was suggested that the E-type pronoun be replaced by an appropriate definite description derived from the antecedent, such as "the sheep John owns". In (145), the description for *he* would be "the boy who deserved it", which can be further expanded to "the boy who deserved the prize he got". Although there is still an anaphor in the description, it can be taken as being bound by "the boy". In effect what we have done is to create an entity for each antecedent, and to constrain them to be related in a number of ways: calling the entities p and b, that b got p, that b deserved p, and that b wanted p.

2.5.4 Alternatives to identity of reference

The data considered so far have all been examples of what Hirst (1981) calls *identity of reference anaphors* (IRAs). That is, they denote the same entity as their antecedent. The theories described in chapter 3, and my own work in chapter 4 are concerned almost entirely with IRAs. Here I will briefly look at some other classes of anaphor.

Hirst distinguishes IRAs from *identity of sense anaphors* (ISAs), illustrated by the "paycheque" sentence:[8]

(146) The man who gave his paycheque to his wife was wiser than the man who gave it to his mistress.

In (146), *it* does not refer to the same paycheque as the antecedent, but to an object which can be described in the same way, namely "his paycheque" or "the paycheque belonging to the man in question". Semantically, we could treat this as a copying of the antecedent's sense rather than its reference. A formalisation might be to take the antecedent as contributing a function from men to paycheques and the anaphor as applying this function to a specified man. Anaphors of the sort in (146) have also been called "pronouns of laziness".

Another identity of sense example is

(147) Neurotics build castles in the air. Psychotics live in them.

Two possible paraphrases of the second sentence are:

(147a) Psychotics live in castles in the air.

(147b) Psychotics live in the castles in the air which (the) neurotics build.

The identity of sense paraphrase (147a) is possible; (147b), in which there is identity of reference, is harder.

Anaphora in modal and temporal contexts may behave in a similar way, as examples (148)–(150) from Cooper (1979), show.

(148) This year the president is a Republican. Next year he will be a Democrat.

(149) The secretary has typed the report himself for years. In the future a typist will often type it for him.

(150) Mary hasn't yet found the man she will marry. Now that she has a job in Washington she hopes that she will meet him.

In none of these cases does the anaphor refer to the same entity as the antecedent, but to some corresponding individual under a different temporal or modal index.[9]

Finally, Hirst lists a wide range of other sorts of anaphora, involving anaphors which need not be pronouns (or indeed NPs), and which stand for sentences, verbs, actions, adjectives, prototypical individuals ("one" anaphora), temporal and physical locations, and verb phrases (VP ellipsis). I shall have nothing more to say about these. "One" anaphora and VP ellipsis are examined in some detail by Webber (1979, 1983).

[8] Attributed by Hirst (1981) to Wasow (1975).

[9] See also the digression at the end of section 2.1.

2.6 Distributive and collective readings

The concepts of distributive and collective readings, and the related issues of covers and partitions, are not directly part of the semantics of anaphora. However, they are needed in order to understand plural noun phrase anaphora, as explored in the following chapters.

The sentence

(151) Four men lifted three tables.

can be read in a number of different ways. It can be understood as meaning that each of four men lifted three tables, that four men together lifted three tables, and that there are a number of collections of men, each collection lifting three tables and with the total number of men being four. The distinction between the three readings is made by Scha (1984), who uses the term *distributive* for the first reading, and *collective* for the other two. Verkuyl and van der Does (1991), in their lucid exposition of Scha's work, label the distributive reading D, and the two collective readings C_1 and C_2 respectively. The same three-way distinction can be applied to the object NP. On the reading where the subject is distributive, the sentence could mean that each man picked up three tables in separate lifting events, that each man picked up three tables in one go, and that each man picked up some piles of tables totalling three, corresponding to the D, C_1 and C_2 readings of the object. The C_2 reading is the hardest to understand. Verkuyl and van der Does clarify the $D/C_1/C_2$ distinction by a diagram illustrating some possible situations which verify (151). Limiting attention to the case where the object is taken as having the C_1 reading, the diagram looks like this, where the brackets enclose men that acted together, and tables that are lifted together:

(152) D reading of subject

$\{m_1\}$ lifted $\{t_1, t_2, t_3\}$
$\{m_2\}$ lifted $\{t_4, t_5, t_6\}$
$\{m_3\}$ lifted $\{t_7, t_8, t_9\}$
$\{m_4\}$ lifted $\{t_{10}, t_{11}, t_{12}\}$

C_1 reading of subject

$\{m_1, m_2, m_3, m_4\}$ lifted $\{t_1, t_2, t_3\}$

C_2 reading of subject

$\{m_1, m_2\}$ lifted $\{t_1, t_2, t_3\}$
$\{m_3\}$ lifted $\{t_4, t_5, t_6\}$
$\{m_4\}$ lifted $\{t_7, t_8, t_9\}$

The number of readings is further increased by scope ambiguities. Scha identifies one additional reading, which he terms *cumulative*, according to which the number of men who lift a table is four and the number of tables lifted by a man is three, but the relation between the men and the tables is unknown. This is often held to be the same as the reading which is C_1 in both subject and object, C_1C_1 in Verkuyl and van der Does' notation. Scha (1991) claims this to be a widespread misunderstanding, and quotes two sentences which show the difference. (153), which was his original example in Scha (1984), can be read cumulatively, but (154) does not allow the cumulative paraphrase in (154a).

(153) 600 Dutch firms have 5000 American computers.

(154) Less than five boys date more than six girls.

(154a) The number of boys who date a girl is less than five, and the number of girls who are dated by a boy is more than six.

The difference can also be illustrated by a diagram for a cumulative case of (151):

(152a) $\{m_1\}$ lifted $\{t_1\}$
$\{m_2\}$ lifted $\{t_2\}$
$\{m_3, m_4\}$ lifted $\{t_3\}$

which cannot be fitted into any of the forms shown above, all of which require that the total number of things lifted by any entity on the left hand side is three. It is not the same as the C_1 reading of both subject and object, as is often claimed, although it does converge with it in the extreme case. For example, in (153), the processing complexity of establishing the exact relationships between firms and computers may be too great, and the interpretation may be simply that a collection of 600 firms considered as a single entity has the 5000 computers considered in the same way. There is some variation in terminology across the literature. Webber (1979) has a collective reading of (151) corresponding to the subject C_1 reading, and a conjunctive reading, which is Scha's cumulative. Kempson and Cormack (1981) use the term incomplete group interpretation for the latter case.

Scha obtains the different readings by making NPs ambiguous between D, C_1 and C_2. For the cumulative reading, he analyses the sentence as a quantification over pairs satisfying the verb, such that the set of such pairs meets the cardinality conditions of the determiners. When scope ambiguity is taken into account, (151) has a total of 19 readings: three readings of each of two determiners, two scopings of each such reading, and the cumulative reading.

A different approach has been taken by Link (1984b, 1987), who obtains different readings from properties of VPs. Link's semantics obtains the ambiguity by applying operators to verb and verb phrase extensions, which either force them to apply to individuals only, for the distributive case, or permit them to apply to collections of individuals. The result is eight readings for (151): two for each argument place of the verb, and two scopings. Of these readings, there are two pairs which coincide, leaving six readings. Link (1984b) explicitly chooses not to consider the cumulative readings, considering it as an extreme example of the reading which is collective on both arguments. He also rejects the existence of C_2, which he calls the *partitional* reading. The reason behind this may be that he does not consider examples which are sufficiently rich. His discussion of this point is based on two examples:

(155) Six boys gather.

(156) Half a million people gathered throughout the country.

In both cases, there can be more than one gathering, provided the total number of individuals involved in the gatherings is as given: the C_2 reading. The C_1 reading is a special case where there is just one gathering, and hence one partitioning of the boys or people. Link initially proposes that an operator could be applied to predicates to derive the C_2 reading, just as he uses an operator for the distributive reading, but subsequently comments that it is preferable not to distinguish the C_1 and C_2 readings, and drops the

C_2 case. In the examples above, this would require that the extension of *gather* includes collections of individuals who gather, whether in one gathering or several.

In discussing his rejection of the cumulative and C_2 readings as special cases of C_1, he raises an important issue:

> Here we touch on a methodological point of a quite general nature in linguistics: where exactly does the line of demarcation run between proper readings and mere models realising a reading? (Link, 1984b, p.23)

That is, if the distinction between the readings can be made by examining properties of the situation being described – for example by counting the number of tables involved in (151) and (152) – why build it into the semantics? The discussion of readings and models realising a reading is taken further by Verkuyl and van der Does (1991). They point out that both the D and C_1 readings can be treated as special cases of the C_2 analysis, for D by requiring the NP be subdivided into individuals, and for C_1 by forcing only one subdivision. They give an illustration by considering a sentence with two NPs, and, for a situation which verifies a given reading, enumerating the set of other readings verified by that same situation. The conclusion is whatever the reading, C_2C_2 appears in the set. It is, in some sense, weaker than any of the other readings.

If this view is accepted, the range of readings for (151) can be reduced to at most two cases: the reading which is C_2 in both subject and object, and the cumulative reading when available. The question is then, in the terms of Kempson and Cormack (1981), whether an ambiguity account should be adopted, or whether a vagueness one is more appropriate. In the former case, all of the readings are made available in the semantics; in the latter, only the weakest reading. Verkuyl and van der Does propose that without additional information, no distinction can be made between the collective and distributive reading, and that a vagueness account (as well as being simpler) is more appropriate. As a first exploration of where additional information might come from, they suggest looking to distinctions based on temporal information. For example, it is easier to distinguish the readings in (157) than in (158), since the tense of the verb allows an appeal to when the events happened.

(157) Three men came into the shop today.

(158) Three men come into the shop.

It is the different events which distinguish the "parts" of the reading, that is, how the referent of the NP is divided up. A precise account of exactly how temporal information and event structure applies to distributivity and collectivity is yet to be developed.

Lexical information will sometimes restrict which readings are available. There are some determiners, such as *most* and *each*, which do not seem to allow a collective reading, and some verbs, such as *gather* and *disperse* which require one. Phrases like *between them*, *together* and *each* may also force certain readings. To illustrate: in (159), *most* forces a distributive reading and *between them* a collective one, and the conflicting constraints render the sentence unacceptable.

(159) ?Most men lifted three tables between them.

Similarly, the distributivity caused by *each* in (160) and (161) is not acceptable with the collective only verbs:

(160) ?Each man dispersed.

(161) ?Six boys each gather.

The classification of determiners as forcing the distributive reading or allowing the collective one is not hard and fast. Some determiners do unequivocally place a NP in one class or another. For example, NPs with *every* are always distributive. It is not always clear, however. Lønning (1987) gives examples which sometimes allow collective readings (in decreasing order of ease): *many/at least three/few/at most four/most/more than half the/at most half the men. Some* appears to allow a collective reading, as in

(162) Some boys buy five roses (between them).

No may also combine with collective verbs:

(163) No men gather.

It is not clear that it is meaningful to say that this is a collective reading, however: is no men doing something as individuals any different from no men doing it together? There do not appear to be any simple determiners in English (or Dutch, according to Verkuyl and van der Does) which force the collective reading, and exclude the distributive one.

Some theories block examples like (159) and (160) by marking them as ungrammatical on the basis of lexical properties of the determiners and verbs. Scha (1984) lets them through, and treats them as being "semantical anomalies" in the same way as

(164) Colourless green ideas sleep furiously.

I find this rather unsatisfactory, although it is hard to produce a cast-iron argument against it. A possible line of attack is that (164) may be made acceptable by changing the context of interpretation (such as treating it as a line of poetry), so that the exact meaning of the lexical items changes, while properties such as singularity and plurality, collectivity and distributivity are kept intact. It is harder to construct such a variation for (160) and (161). A model which could make (164) true can be imagined; for (160), it is difficult to do so without treating either *man* or *disperse* as meaning something radically different from its conventional use.

If lexical information sometimes blocks the D reading or the collective ones, it follows to ask whether C_1 can be blocked and C_2 allowed, or vice-versa. There is no compelling evidence to show that this is so. Even modifiers like *together* and *between them* do not force one collective reading over the other. The availability of C_2 does seem to vary with models and with listeners. I find the C_2 reading of (151) possible, but barely so. Verkuyl and van der Does report that in a situation which verifies only the C_2 reading, such as

(165) $\{m_1, m_2\}$ lifted $\{t_1, t_2, t_3\}$
$\{m_3, m_4\}$ lifted $\{t_4, t_5, t_6\}$

one of them finds it possible to so read the sentence and the other does not, and I have had similar reactions from informants. Contextual factors such as background knowledge may make a difference. Consider

(166) Four millionaires share three villas.

in a situation where, as a result of shortage of luxury accommodation, all millionaires have been summoned to prove that they share their dwellings. Two of the millionaires are examined and prove that they have three villas between them, and then two further ones are so examined and prove the same thing, although their villas are different from the first three. This is certainly a C_2 reading – it is verified by a situation with the same structure as (165) – and I find it perfectly acceptable, given the context.

Some confusion arises with sentences like (167).

(167) Most sheep flock.

The determiner forces distributivity, but the verb requires a collective reading: a single thing cannot flock. Nevertheless, (167) is not as clearly unacceptable as (159) and (160). The sentence becomes less acceptable when the verb tense directs attention towards some particular event or situation and away from the generalisation.

(168) ?Most sheep are flocking.

An explanation is that the interpretation of the verb has to change with its tense, perhaps in keeping with (157)–(158). In (167) it is interpreted as holding of all things that, in general, tend to go round in flocks; in (168), it holds of things that form a flock at the instant specified by the tense, and hence is true only of collections of sheep and not individual ones.

The grammar constructed by Verkuyl and van der Does does not rule out any combination of NPs and verbs on the basis of syntactic features: (159) and (160) are parsable. The semantics yields the C_2 reading in all cases. For a NP that must be distributive, the distributivity is built into the semantics of the determiner, by discarding plural objects from the sets to which they are applied. (159) and (160) then become unacceptable because there are no models which realise them. Determiners which allow both distributive and collective readings are treated uniformly, i.e. as vague between the interpretations. To block (160), Verkuyl and van der Does use a meaning postulate on the extension of *disperse*, requiring that it consist only of plural objects. The definition of *each* removes any plural objects from the extensions, and hence (160) can never be satisfied. The fact that the sentence is false in all models can be deduced from the interpretation of (160) without knowing exactly what the extension of *disperse* is in any particular model.

Two final points. Firstly, there may be yet further readings. In van der Does (1992)[10] there is a proposal that there may be a second distributive reading and two more collective ones, distinguished from D, C_1 and C_2 by a notion of participation, so that each individual or collection of individuals does not have to do what is predicated of the NP, but only to play some role in doing it. If this proposal is adopted, it is again possible to find one reading of which all the others are special cases, but it hardly ever occurs as a natural reading. "Participatory" readings may provide the solution to example (167). I will not explore this issue further, and restrict attention to the three-way $D/C_1/C_2$ split.

Secondly, with singular NPs, there is no need to make any distributive/collective distinction. One man lifting a table as an individual is the same as one man lifting a table as a collection. This leads Lønning (1987) to divide NPs into three classes: singular, plural and quantificational. Quantificational NPs are those which admit only a distributive reading. The singular classification does not correspond exactly to the syntactic one. For example,

[10]Also published, in a slightly shorter form, as van der Does (1991). For some related discussion, see Lønning (1991).

every man has the syntactic properties of a singular NP, but is still quantificational. There are some limitations on the kind of verb that singular NPs can combine with. It does not make sense to say

(169) A man gathers.

Examples like (169) could be ruled out syntactically, or we could just say that models with a typical extension for *gather* always falsify it, in a similar way to above.

2.6.1 Covers and partitions

In C_2 readings, it may be necessary to place some restrictions on how the entire collection of individuals described by a NP may be subdivided. For example, in (151), we need to ask whether the groups of men which each lifted three tables might have members in common. Four possibilities are suggested in van der Does (1992, pp.44-55). Related discussions can be found in Verkuyl and van der Does (1991) and Gillon (1987).

1. Cover: all non-empty sub-collections are allowed.
2. Pseudo-partition: the number of sub-collections is no more than the number of individuals in total.
3. Minimal cover: a cover containing no subset which is also a cover.
4. Partition: no two sub-collections may have any members in common.

Each item in the list includes the following items as special cases. There is some evidence against each of the possibilities. A partition will not do for

(170) Hammerstein, Rodgers and Hart wrote musicals.

because the sentence is still judged true if Hammerstein and Rodgers wrote some musicals together and Rodgers and Hart wrote others, but none of them wrote musicals in any other combinations. A minimal cover is wrong for

(171) Three managers got together.

which is true in a situation where all three managers got together and separately two of those same managers got together. The argument for preferring pseudo-partitions over covers derives from

(172) Four men lifted two tables.

which on a cover interpretation permits as many as thirty tables in total to have been lifted: two tables multiplied by fifteen possible subsets of the four men.[11] This seems absurd, and the pseudo-partitional reading attempts to correct it by ensuring than no more than eight (= 4 x 2) tables are lifted, even if some of them are lifted by sub-collections containing more than one man. (171) can be used as an argument against pseudo-partitions. The situation where all three managers got together, as did all sub-collections of two managers, is not a pseudo-partition, but seems a reasonable situation for the sentence to be judged as true. The implication is then that a full cover is allowed, at least in intransitive cases. Conversely, for transitive verbs, pseudo-partitions may not be restrictive enough. In

[11] Van der Does gives thirty-two rather than thirty, by allowing the empty set in the cover.

(173) Richard and Harry each lifted two tables. Richard and Ellen lifted two tables. Three people lifted two tables.

pseudo-partitioning predicts the last sentence to be true, but this goes against the intuitions of a number of semanticists (van der Does cites Lasersohn, Lønning and Verkuyl). As usual, such judgements vary with context. In a situation where it is being checked whether Richard, Harry and Ellen have satisfied their duty of lifting three tables, the last sentence of (173) could well be judged true, as would

(174) Three people lifted two tables alone or with others.

The semantics in Verkuyl and van der Does (1991) adopts pseudo-partitioning as the correct reading of collective predication, but, as van der Does notes, its use is still debatable.

3 Two Theories of Anaphora

In chapter 2, I surveyed the data on noun phrase anaphora, concluding that it is hard to arrive at a definitive statement of what some sentences containing anaphors mean, particularly where quantifiers and conditionals are involved. Prior to the early 1980s most semanticists were concerned with analysing and classifying the data, leading to work such as Evans' categorisations of pronouns. It was assumed that existing semantic frameworks, in particular Montague Semantics could be used to formalise the analysis.

Recent work in the semantics of anaphora has proceeded differently, by adopting new logical formalisms. In this chapter, I look at the two major theories of this sort: Discourse Representation Theory, and dynamic logics. What these theories have in common is that they start from the assumption that a truth-conditional approach to semantics is not rich enough. Discourse Representation Theory, which is most strongly associated with Hans Kamp, makes uses of a level of representation intermediate between the surface form of the discourse and its truth conditions. Unlike the formulae of Montague Grammar, which are a convenient and dispensible technical device, the representations of DRT capture information about the discourse which has no direct counterpart in the model-theoretic interpretation. DRT's first application was to anaphora. Its methodological and formal principles have since been used to study a variety of semantic phenomena. Dynamic logics, of which Groenendijk and Stokhof's DPL is perhaps the best known example, employ syntactically standard logical formalisms with non-standard interpretations, in which the truth conditions are secondary to the change in information state resulting from the process of interpretation. A variety of different dynamic logics have been proposed, the more recent ones converging – in some respects – with Montague Grammar.

DRT is described in section 3.1. The techniques of the theory have been applied in several areas of semantic research, and I briefly survey both the background to the theory and its major non-anaphoric applications, before turning to its specific use in anaphora. One of its more striking aspects is that indefinite NPs have no intrinsic quantificational properties, gaining their interpretation from the presence of other NPs. This view is usually referred to as "indefinites as (free) variables", and leads to one particular interpretation of the donkey data. In section 3.2, I turn to dynamic logics. Groenendijk and Stokhof have claimed that their work is non-representational, placing it more in a Montagovian tradition than DRT is. I will test out this claim, leading to a more precise notion of what it means for a theory to be representational. One source of confusion about the term may be cleared up immediately. The sense in which representation is being used in DRT and by Groenendijk and Stokhof is that of an essential representation; that is, something which contains information not present in either the discourse or the model with respect to which the representation is interpreted. This is sometimes glossed as the theory having a "level of representation". The term is also used, for lack of something better, simply to mean the notational system used to abstract away from details of the model, such as the use of predicate symbols rather than their extensions, or of logical connectives rather than their meta-language description. It should be clear from context when the "essential" sense is meant.

3.1 Discourse Representation Theory

It is perhaps misleading to treat discourse representation theory as being a single entity. As with Montague Grammar or Situation Semantics, it can be seen as a methodological approach to the formal study of natural language, which is used to guide the semantic analysis of specific natural language phenomena. In this section, I will explore the central ideas of discourse representation theory, and then turn to applications of it.

Discourse representation theory is most closely associated with the theory proposed by Hans Kamp and work stemming from it, generally referred to as DRT. The File Change Semantics of Heim (1982, 1983) has also been described as a discourse representation theory. Heim's work addresses similar semantic problems to Kamp's early work in DRT, namely anaphora and definite and indefinite noun phrases, and makes similar empirical predictions. There has been no significant development of File Change Semantics beyond Heim's initial statement of it, and I will concentrate on Kamp's DRT here.

3.1.1 The philosophical background to DRT

The aim of DRT, as summarised in Kamp (1981) (reprinted as Kamp (1984)), is to unify two notions of natural language meaning. The first notion treats meaning as that which determines truth conditions. Truth conditional semantics is typically model-theoretic, in the sense that the meaning of a discourse is treated as the conditions it would impose on a model of possible or actual worlds in order for the discourse to be judged as true by a speaker of the language. The second notion is that meaning is, in Kamp's words, that which a language user grasps when he understands the words he hears or reads. DRT brings together the two notions of meaning by introducing a level of representation intermediate between the surface form of the discourse and the truth conditions. Formally, the intermediate representation is expressed as a collection of discourse representation structures (DRSs). The second notion of meaning is captured in the process which is used to construct the DRSs, and the first in the truth conditions that arise when a relation is established between DRSs and models. DRSs can be seen as partial models, each containing information about the discourse that is coherent but incomplete. The truth conditional interpretation of DRSs consists of embedding them in a larger model, which takes its content from an abstraction of an actual or possible world.[1]

DRT may be contrasted with theories such as Montague Grammar, where the representation is inessential. As Dowty et al. (1988) say of Montague's IL:

> translating English into Intensional Logic was therefore not essential to interpreting the English phrases we generated; it was simply a convenient intermediate step in assigning them meanings. (p.263)

Kamp (1984/85) suggests two advantages that DRT has over earlier semantic theories. Studies carried out in the sixties and early seventies – Kamp cites Montague, Scott, Kaplan and Lewis – paid some attention to the role of context in interpreting discourses, both in non-linguistic factors such as the time and place of the utterance and the identity of the participants, and in linguistically realised elements such as tenses and pronouns. Kamp's criticism of this work is that it does not go far enough: it acknowledges that context is needed to interpret an utterance, but ignores how the interpretation of one utterance establishes the context for subsequent ones. That is, not only does context affect the interpretation of utterances, but the converse is also true. Kamp calls this the principle of the unity of the context and the content.

[1]This should not be taken to indicate that there is a necessary commitment to a model-theoretic approach in DRT. However, the formal aspects of DRT have generally been expressed in model-theoretic terms.

The second claim in Kamp (1984/85) is that the structural constraints on DRSs required to describe the meaning and cohesion of discourse also contribute to a more general understanding of linguistic communication. In particular, he proposes that a cognitive state can be identified with a family or network of DRSs, each capturing some aspect of cognitive structure (p.247). For example, we might choose to associate a DRS with each belief held by an individual, possibly allowing such things as incomplete and inconsistent belief states.

DRT's role as a theory of verbal communication and cognitive representation is developed further in Kamp (1991a). Cognitive states are now identified with DRSs which contain a mode indicator representing the relation of the cognitive agent to the thought represented. Kamp proposes that the process of verbal communication can be identified with the process of constructing DRSs. Speaking consists of the conversion of a thought in the form of a DRS in the speaker's mind to a sequence of words, and hearing to the process of converting the words into a DRS in the hearer's mind. The hearer's DRS need not be identical to the speaker's: it may contain (for example) different mode indicators, and extra information describing the source of the information and the hearer's beliefs about the speaker's attitude towards the thought. However, if the communication is successful, some parallel between parts of the speaker's and hearer's DRSs will exist.

Guenthner (1987) examines the role of DRT in a general theory of discourse meaning. He divides meaning into three components: linguistic meaning, truth conditions and denotation conditions. Linguistic meaning consists of how representations are constructed from discourses, truth conditions of the procedure for embedding a DRS in a model or set of models, and denotation conditions of the relation of models to some otherwise given information. DRT provides a means of studying the first two components. Guenthner also proposes that properties of DRSs could be investigated by other means than their truth conditions. Firstly, a deductive system of inference could be formulated, just as there are model-theoretic and proof-theoretic approaches to many systems of logic. Secondly, DRSs could be investigated from a psychological perspective. For example, DRSs might be considered as an approximation of mental models in the sense of Johnson-Laird (1983). Finally, DRSs can be treated as being similar to "deductive databases" used in some theories of logically based information processing; a computational use of DRSs. There is thus some possibility of drawing together insights about discourse interpretation from a variety of sources: mathematical, psychological and computational. Because the form of DRSs is stated independently of which manner of investigation we adopt, Guenthner claims that the empirical load of DRT should be located in the process that constructs DRSs from discourses. Kamp (1984/85) presents a similar view, when he says that "from a linguistic standpoint the real challenge to [DRT] is to come up with an exact and concise statement of [DRS construction] rules. (p.242)"

A final idea in Guenthner's paper is that DRT makes it possible to represent two sorts of semantic relation, which he calls D(iscourse)-relations and T(ruth)-relations. D-relations are relations between DRSs, and in particular how DRSs are extended as the interpretation of a discourse progresses. T-relations are relations between DRSs and models. Guenthner's suggestion is that some semantic phenomena are best investigated through D-relations and some through T-relations. The former category includes presupposition, ambiguity, and coherence; the latter truth, consistency and vagueness.

If the claims of Kamp and Guenthner are accepted, one important question arises: when should a semantic phenomenon influence the form of the DRSs (D-relation), and when is it better considered as part of the interpretation stage (T-relation)? Empirical evidence may not always lead to a definitive answer to this question, and there are methodological arguments in favour of both standpoints. By keeping DRSs simple, we avoid obscuring insights into verbal communication and the representation of thought. By keeping the truth definition and the model structure simple, we can better understand notions of information and knowledge, in isolation from linguistic structure.

In the next section, I give a general overview of the formalisms used in DRT, and then briefly look at some of the semantic phenomena to which DRT has been applied. The most extensive work in DRT has been in the study of anaphora, and this forms the subject of section 3.1.4. Despite its successes in this area, there are a number of problems with DRT's account. Some of its empirical predictions are incorrect, or at least over-constrained. There are some difficulties in the interaction of DRT with other semantic theories. Finally, some aspects of DRT applied to plural anaphora do not rest well with the philosophical claims of Kamp and Guenthner.[2]

3.1.2 DRS languages

A number of formal languages of discourse representation structures have been developed. In a typical DRS language, a discourse representation structure K is a pair $\langle U, Con\rangle$, where U is the universe of K, and Con the conditions of K. The universe consists of a set of *discourse referents*, whose properties are determined by the conditions. In anaphoric DRT, discourse referents stand for individuals or collections of individuals. Other DRS languages have used discourse referents that stand for (amongst other things) predicates, beliefs, concepts, events and times. A "box" notation is generally used for DRSs, with the top line listing the discourse referents and the remainder of the box containing the conditions. Examples appear in subsequent sections.

Conditions are used to express constraints on discourse referents and relations between them. Two main kinds of conditions are used: predicate symbols applied to one or more discourse referents, and operations on *sub-DRSs*, that is DRSs contained within other DRSs. The latter allows the discourse described by a DRS to be subdivided, in such a way that properties of the discourse represented in a sub-DRS may be considered independently from the DRS containing it, and the internal structure of the sub-DRS possibly hidden from the rest of the representation. Operations on sub-DRSs include logical connectives, modal operators, quantification, and abstraction of discourse referents from a sub-DRS.

DRSs are built up from a discourse using a construction algorithm. Several different ways of stating construction algorithms have been used in the literature. The clearest is probably that of Kamp and Reyle (1990), where it takes the form of a collection of rules, each consisting of a triggering condition and the actions to be taken when the condition is satisfied. The triggering condition is expressed as a (partial) parse tree. The actions introduce new discourse referents and conditions, possibly including sub-DRSs. The construction starts by creating a "topmost" DRS into which the entire discourse, in the form of a collection of parse trees, is entered. As the rules are applied, the parse trees which have been matched are erased or replaced by simpler trees, until the entire discourse has been processed or the algorithm blocks. This approach differs from many semantic theories in being top-down and non-compositional: it starts from the whole parse tree

[2]The term *DRT* is often equated with the DRT treatment of anaphora, particularly in the form described in Kamp (1981). As I said above, I am using DRT to mean the whole of the discourse representation theory deriving from Kamp's ideas. Where it may not be clear from the context, I will refer to the DRT applied specifically to anaphora and quantification as "anaphoric DRT".

and progressively rewrites it. A compositional, bottom-up reconstruction of the version of singular anaphoric DRT in Kamp (1981) has been developed by Zeevat (1989) (also reprinted in Zeevat (1991)).

A DRS language is given a truth definition in the form of verification conditions recursively defined on the expressions of the language. Typically, the verification conditions are expressed in terms of an *embedding function* (also called an *anchoring*) from a DRS K into a model M, which assigns values to the discourse referents of K, such that the conditions of K are satisfied. The verification conditions of operations involving sub-DRSs are typically expressed as constraints on the possible embeddings of the sub-DRS. The range of possible embeddings of a sub-DRS K' cannot usually be considered in isolation from the DRS within which it occurs, since conditions of K' may contain discourse referents of higher DRSs. Verification conditions are therefore often expressed in terms of extending one embedding to another, by adding bindings for the discourse referents of the sub-DRS to ones for the higher DRSs. Embeddings are similar to the assignment functions of some model-theoretic logics.

A final important idea is accessibility, which specifies which discourse referents are available for use in DRS conditions. An example is the anaphoric accessibility relation, which specifies when an antecedent's discourse referent is available for use in the translation of an anaphor. Accessibility relations are expressed in terms of the DRS configuration.

In the descriptions of the theories that follow, I will make some reference to a DRS language being "standard" to a greater or lesser degree. By this I mean it resembles or extends the language of Kamp (1981). Most DRSs languages follow the standard form. One exception is the treatment of tense in Kamp and Rohrer (1983), which uses the ideas but not the formalism of DRT.

3.1.3 A survey of non-anaphoric DRT applications

3.1.3.1 Attitude reports

The study of attitude reports is of some importance, in that it lies on the boundary between the philosophical claims about DRT as a theory of verbal communication or cognitive representation, and use of DRT in the logical analysis of specific linguistic phenomena.

Zeevat (1986) presents a theory of belief sentences in DRT as part of a wider discussion of a phenomenological theory of thought, according to which the objects that are studied are internal to the the cognitive agent holding the belief. He contrasts this approach with a realist theory, in which the objects studied are considered to have a real existence, and so may be studied independently of the cognitive agent. The phenomenological approach is adopted by Zeevat for two reasons:

> First [...] there seems to be no interpretation of DRT as a realist theory of propositions [...] (Footnote: It would be necessary in order to find a realist theory that follows DRT to give a version of DRS's that does not use variables any more. I doubt whether there exists such a version.) Secondly, there are problems that cannot be handled by a realist theory (or at least lead to a considerable uneasiness) but can be handled in DRT. (pp.189-190)

Zeevat argues that the syntax of DRSs makes it possible to interpret Meinongian objects, i.e. objects which have an internal role. The roles may be derived from structural properties of DRSs: for example, an object is judged to be *existent* if it is a discourse referent of a DRS judged to be true, or *arbitrary* if it appears in the left hand side of a conditional.

Other role are derived from logical properties of the DRS: an object is *necessary* if it is described by a logically true DRS and *possible* if described by a contingently true DRS. Finally, the role may come from the treatment of a DRS as representing a mental state or thought: hence objects of belief or desire, fictional objects, and so on. Most of the paper concentrates on objects of belief, following the view that "the belief state of a person is structurally similar to a large DRS (p.197)". The DRS language is fairly standard, the main change being the introduction of a belief operator which relates a discourse referent to a DRS believed by the individual for which the discourse referent stands. The verification conditions use an embedding function which takes discourse referents to either objects in the model, or to an internal object consisting of an individual and an object. A belief held by an individual is true if the corresponding DRS appears in the belief state of the individual, possibly under a relabelling of objects appearing in the DRS to their private counterparts. Zeevat illustrates this approach with a variety of examples including *de re/de dicto* beliefs, Geach's Hob Nob sentences, and substitution of identical objects into a belief state (Kripke's Pierre story).

Asher (1986) has also developed a theory of belief reports in DRT. His main goal is to overcome some of the limitations of the theories of Hintikka, Montague and others. In particular, these theories permit the substitution in belief contexts of logically equivalent expressions and of coreferential directly referential expressions, neither of which is valid. Like Zeevat, Asher's theory represents belief states as DRSs. DRSs, or rather delineated sequences of DRSs, are used to

> furnish something like a structural description identifying the cognitive state (or some portion thereof) of the believer, including the Belief that is the target of a belief report. (p.166)

Roughly speaking this means that both the content and structure of DRSs form part of the description of the belief, leading to a much more fine grained definition of logical equivalence than is possible in a theory like Montague Grammar, which treats the structure of the semantic representation as arbitrary. Two beliefs are identical only if they involve the same notional objects, standing in the same notional relations, and having the same notional properties, where "notional" is to be understood as forming part of the mental world of the believer.

The DRS language is more complex than Zeevat's, with discourse referents standing for beliefs, DRSs and concepts. Beliefs are represented by predicates which take an individual discourse referent and a belief discourse referent as arguments. Conditions link belief discourse referents to DRS discourse referents, which are in turn equated with DRSs. Concept reference markers identify notional individuals, by means of conditions which link them to individual discourse referents, and conditions which indicate when concepts are to be distinct. The latter point allows inconsistent beliefs to be maintained. The inconsistency only becomes apparent if some extra condition, such as an external constraint, requires the beliefs to be identified with one another

Asher continues his investigation of attitude reports in Asher (1987), by examining the anaphoric properties of a range of attitude verbs. He proposes eleven classes of attitude verbs, and lists which logical inference rules hold valid for each class. Asher states his main objective of this classification as being

> to study the anaphoric properties of a wide range of attitude constructions and to see how anaphoric properties correlate with the logic of the attitude

constructions. A DR theoretic approach predicts that there should be a close correlation; I will show this prediction is borne out. (p.125)

An example of this is the effect of positive indefinite non-factives, such as *believe*, in blocking anaphora. Following

(1) John believes a woman broke into his apartment.

it is not acceptable to say

(1a) She is now in hiding.

but anaphora is allowed if the belief context is re-entered:

(1b) He believes she is now in hiding.

This may be contrasted with positive indefinite factives, for example *be aware*, which permit the second sentence. Such properties are investigated at length for several of the classes proposed by Asher. The paper may appear to be an example of a cross-over from DRT applied in one area (attitude reports) to DRT applied in another (anaphora and quantification). However, the connection is largely superficial. Asher deals with only co-referential anaphora, modified by the attitude context, rather than any of the more complex uses of anaphora which motivated Kamp's original work on DRT. The only similarity with anaphoric DRT is that accessibility relations are used to determine when co-reference is allowed. As discussed in section 2.4, accessibility may not be the only way of restricting anaphoric relations under quantification, and it is possible arguments for other approaches could be made for anaphora in attitude contexts.

Zeevat's and Asher's theories of attitudes can be counted as a significant success for DRT, in that they advance solutions to long standing problems in philosophy of language and natural language semantics. They also provide some support for Kamp's claims about DRT as a theory of cognitive representation, for example, in Zeevat's discussion of phenomenalism and realism.

3.1.3.2 Other applications

A theory which makes use of D-relations between DRSs, in Guenthner's sense, is the treatment of presuppositions by van der Sandt (1990). He examines examples like:

(2) If John made coffee, his wife will be happy.

which presupposes that John has a wife. (2) may be contrasted with

(3) If John is married, his wife will be happy.

where the presupposition is not needed, since *married* tells us directly that John has a wife. Van der Sandt suggests that some problems of presupposition projection may be solved using an accessibility mechanism of the sort that DRT uses for anaphora. For example, he notes that there is a parallel between the anaphoric (4) and the presuppositional (5).

(4) If John owns a donkey, he beats it.

(5) If Jack has children, all of Jack's children are bald.

Following (4), anaphora to *a donkey* is not permitted; following (5), the presupposition is no longer held. Other theories of presupposition have used notions such as cancelling of implicatures when the presupposition is no longer needed. Van der Sandt proposes that accessibility provides an alternative solution. He represents presuppositions as conditions appearing in the DRSs associated with certain lexical and phrasal items. To accommodate the presuppositional conditions, they are matched against conditions already present in DRSs accessible from the one where they appear. Any condition for which no match can be found is accommodated by inserting it into the highest accessible DRS. Van der Sandt summarises the main claim of his paper as:

> Presuppositions are simply anaphors. The only differ from pronouns or other kinds [of] semantically less loaded anaphors in that they contain enough descriptive context to establish a reference marker in case discourse does not provide one. (p.4)

Van der Sandt's theory supports the view that there is a need to structure the analysis of the discourse, using accessibility to determine the scope of presuppositions. His theory is a good illustration of how the D-relations may be useful as a technical device, whether or not a commitment is made to DRSs as cognitive representations.

The same is true of Asher (1990), which examines three problematic uses of gerunds using a DRS language for the semantics. Asher comments that DRT allows one to

> make these distinction between abstract objects at the level of natural language metaphysics without being committed to the same distinctions in the model theory.

Finally, there are several theories which use DRT but which are of minor importance for the present discussion. Klein (1986) uses a DRS language with predicative DRSs – lambda-abstractions of a variable from a DRS – for a treatment of VP ellipsis. Zwarts (1990) gives a treatment of generics based on predicative DRSs and a "generic quantifier". Partee (1984) exploits the parallels between nominal and temporal anaphora to use DRT in a representation of tense and temporal reference, with discourse referents representing events and times. DRT's ability to express subordinating and coordinating relations between events by means of relations between sub-DRSs has been exploited by Lascarides and Asher (1991), who are concerned with defeasible entailment in discourse. Bartsch (1987) represents "frames" as DRSs, with the frames being used for a variety of purposes including lexical constraints ("a bachelor is an unmarried man"), prototypical scenes ("restaurant visit") and constitutive relations ("walls are part of a house"). Stirling (1988) examines languages where there are markers which constrain anaphoric relations, as reflexivization does in English, and uses an extended version of the standard DRS language as the semantic representation.

3.1.4 DRT and anaphora

The paper in which Kamp introduced DRT (Kamp, 1981) describes a theory of noun phrase anaphora. The coverage of the original version is very limited: only singular pronominal anaphors are considered, and the only antecedents are proper nouns and noun phrases with the determiners *every* and *a*. Much work has been done to develop the theory since then. Kadmon (1985) adds a simple treatment of numeral determiners. Sells (1985) and Roberts (1987, 1989) also introduce some changes, arising from the issues of non-restrictive modification and modal subordination respectively. More recently, there has been an extensive revision of DRT to cover plural anaphora (Kamp and Reyle, 1990).

DRT assumes that the antecedent for each anaphor is known. A discussion of how properties of DRSs could be used to guide selection of an appropriate antecedent (anaphor resolution) can be found in Guenthner and Lehmann (1983). Kamp's theory and most of the modifications of it assume that antecedents textually precede anaphors; cataphora are discussed by van Deemter (1990).

A number of computational implementations of the DRT treatment of singular anaphora have been developed: see Frey and Reyle (1983), Johnson and Klein (1986), Reyle (1985), Sedogbo (1988) and Wada and Asher (1986).

In this section, I will concentrate on Kamp's treatments of singular and plural anaphora, presenting the theory in the style of Kamp and Reyle (1990). The DRS languages and their construction algorithms and verification conditions are listed in full in appendix A; details are given here as necessary.

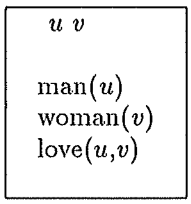
3.1.4.1 DRT and singular anaphora

The DRS language used by Kamp for his treatment of singular anaphora follows the standard form described in section 3.1.2, with discourse referents standing for individuals. There are three sorts of condition: n-ary predicate symbols applied to n discourse referents, equality between two discourse referents and *implicative* conditions, appearing as a logical connective between two DRSs.

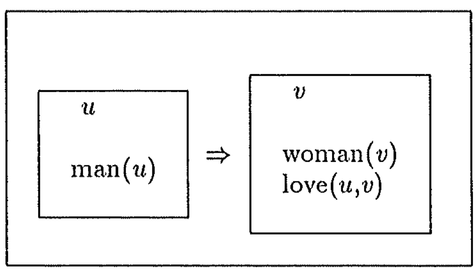
The fragment of English includes singular pronouns and proper nouns, singular noun phrases with determiners *every* and *a*, and intransitive and transitive verbs. The determiners are introduced syncategorematically, as is a conditional construction *if* $sentence_1$, $sentence_2$.

By way of examples, the DRSs for (6) and (7) are as shown below.

(6) A man loves a woman.



(7) Every man loves a woman.



In the DRS for (6), u and v are the discourse referents for *a man* and *a woman* respectively. The two sub-DRSs in (7) are an example of an implicative condition, triggered by *every*.

All NPs, including pronouns, introduce new discourse referents into the universe of their DRS. Where the NP has determiner *every*, an implicative condition is introduced with the nominal appearing in the left hand sub-DRS and the body – in (7), the VP – in the right-hand sub-DRS. The condition for pronouns equates their discourse referent with one from another NP, under the restriction that it must be suitable and accessible. Suitability means that syntactic constraints such as (in English) gender are satisfied. Accessibility is defined in terms of structural relations between DRSs. A discourse referent u is accessible from a DRS K if either $u \in U.K$ (where $U.K$ is the universe of DRS K); or K appears in a condition of the form $K_1 \Rightarrow K$, and $u \in U.K_1$; or K is a sub-DRS of a DRS K_1, and u is accessible from K_1. For example, if the universe of the DRS in (6) contained another discourse referent w, then both u and v would be accessible to it, as needed for

(8) He/She is happy.

However, in (7), neither u nor v are accessible to discourse referents introduced into the topmost DRS, in keeping with

(9) Every man loves a woman. ?He/She is happy.

When there is no suitable and accessible antecedent discourse referent for an anaphor, it is judged as unacceptable.

Proper nouns are treated similarly to ordinary NPs, except that the discourse referent is always introduced into the top-most DRS, meaning that it will be accessible regardless of where it appears in the surface form. Consequently, (10) is acceptable, where (11) is not.

(10) Every man likes Sheila. She is happy.

(11) Every woman likes a man. ?He is happy.

The approach of putting proper nouns at top-level is quite different from most theories of anaphora, which treat them no differently to other antecedents, and relies on the construction process working top-down, so that the topmost DRS is known.

The truth definition for DRSs is expressed in terms of embedding functions (or simply embeddings), which map the discourse referents of a DRS to individuals in a model. An embedding is said to *verify* a DRS if all of the conditions can be satisfied. Where the conditions include sub-DRSs, extra bindings are added to the embedding function for the discourse referents of the sub-DRS, *extending* an embedding to the universe of the sub-DRS. An implicative condition $K_1 \Rightarrow K_2$ is verified by an embedding f if for every embedding g which extends f to the universe of K_1 and verifies K_1, there is an embedding h which extends g to the universe of K_2 and which verifies K_2. In effect, this means that all discourse referents in the universe of K_1 are universally quantified. The accessibility relation can be incorporated into the verification definition, as Chierchia and Rooth (1984) show: if embedding functions are partial, and a discourse referent x is inaccessible, then $f(x)$ will be undefined, and the attempt to verify the DRS condition containing x will block. However, this is still a structural definition of accessibility and not a semantic one, in the terms of section 2.4.

Heim (1990, p.137) summarises four main claims made by anaphoric DRT and her own File Change Semantics (Heim, 1982), in contrast to earlier theories of anaphora:

1. Indefinites are not quantifiers, existential or otherwise.
2. Anaphoric pronouns are treated as bound variables, rather than definite descriptions.
3. Quantifying determiners and the conditional operator can bind multiple variables.
4. Variables not bound by a quantifying determiner or the conditional operator are given an existential interpretation.

The first point arises from the construction rule for *a*, which introduces a new discourse referent but not a new sub-DRS. The DRS which contains it will be either one created by the rules for *every* and *if*, or the top-level one, Heim's third and fourth points respectively. A consequence of Kamp's approach is that bound, co-referential and E-type anaphors receive a uniform treatment. Note also that the four claims depend on analysing anaphora as a D-relation: it is the structure of DRSs that gives indefinites and anaphors their interpretation.

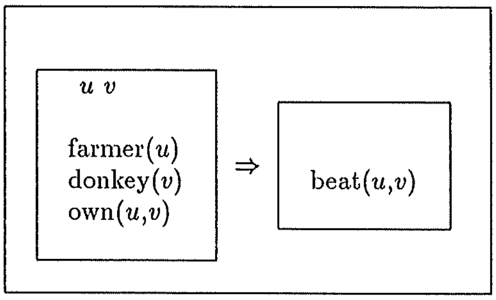
The claims can be illustrated by reference to three variants of the donkey sentence:

(12) Every farmer who owns a donkey beats it.

(13) If a farmer owns a donkey, he beats it.

(14) A farmer who owns a donkey beats it.

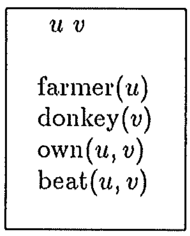
For both (12) and (13), we get the DRS:



From the verification definition, it follows that both u and v are treated as bound by a universal quantifier, and the sentence has a universal reading:

(15) $\forall x \forall y[(farmer(x) \wedge donkey(y) \wedge own(x,y)) \rightarrow beat(x,y)]$

The DRS for (14) is:



with interpretation equivalent to:

(16) $\exists x \exists y[farmer(x) \wedge donkey(y) \wedge own(x,y) \wedge beat(x,y)]$

In this example, neither indefinite noun phrase appears within a universal quantification or conditional, and so they are both read as existential. The accessibility relation blocks the reading of (17) which has *he* anaphoric to *every farmer* and of (18) with *it* anaphoric to *a donkey.*

(17) Every farmer who owns a donkey beats it. ?He is cruel.

(18) Every farmer who owns a donkey beats it. ?It brays in distress.

Heim contrasts the four main claims made by DRT with the alternatives in the theories such as Evans' and Cooper's, namely:

1. Indefinites are existential quantifiers.
2. Anaphoric pronouns are semantically equivalent to (possibly complex) definite descriptions.
3. Quantifying determiners, frequency adverbs, and the hidden operator of generality in conditionals bind just one variable.
4. There is no need for default existential quantification of free variables.

The Kamp-Heim approach is therefore more than just a matter of finding how to solve problems of quantification and anaphora within existing semantic theories. It involves a significant change in how indefinites and anaphors are to be thought about. In section 3.1.5.3, I will examine the necessity for and correctness of this shift.

3.1.4.2 DRT and plural anaphora

The DRT treatment of plural anaphora is described at length in chapter 4 of Kamp and Reyle (1990). I omit the parts of the theory relating to genera, dependent plurals, floating quantifiers and reciprocals, and concentrate on the main part of it: the treatment of plural pronominal anaphors with noun phrase antecedents.

In the syntactic fragment, Kamp and Reyle assign a number feature to noun phrases and determiners, taking the values singular or plural. Determiners also have a "quant" feature, which marks them as quantifying, indefinite, or definite. Examples of the three classes are (quantifying) *every*, *most*, *few*, *no*; (indefinite) *some*, *several*, *a(n)* and a null determiner; (definite) *the.* Numerals may be treated as quantifying, for *at least two/three/...* and *exactly two/three/...*, or as indefinite, for bare numerals.

Discourse referents stand for collections of one or more individuals. Several additional kinds of condition are added to the DRS language.

1. Atomicity and non-atomicity conditions, to force discourse referents to be singular or plural respectively.
2. Membership of a singular discourse referent in a plural one.
3. Cardinality constraints on discourse referents, for numeral determiners.
4. Generalized quantification over a discourse referent, holding between two sub-DRS. Such conditions are called *duplex* conditions.

Two new operators are introduced, one summing two or more discourse referents, and the other abstracting one discourse referent from a DRS and summing over its possible values. A typographical convention is usually adopted as shorthand for the atomicity conditions: lower case letters for atomic discourse referents and upper case for non-atomic ones. Atomic discourse referents are also referred to as individual, and non-atomic ones as plural. In addition, individual discourse referents may be annotated with the symbol *pl*, for reasons discussed below. Sums of discourse referents ($X = y \oplus z$) are used for antecedents formed by combining other antecedents, as in

(19) John saw Mary at the party. They had a good time.

The membership condition ($x \in Y$), used with a duplex condition, allows iteration over the individuals comprising a plural discourse referent.

The remaining two conditions need a little more explanation. Abstraction, $x = \Sigma z K$, forms the sum of all individuals z which satisfy the conditions of the DRS K. It is used in examples like:

(20) Susan has found every book which Bill needs. They are on his desk.

where the discourse referent for *book* is abstracted and summed. The construction rule for abstraction applies during the processing of the antecedent sentence, rather than "on demand" when *they* is processed.

Duplex conditions, $K_1[Qx]K_2$, are used for quantification over a discourse referent, and stand for generalized quantifiers. In generalized quantifier theory, determiners can be expressed as relations between two sets of individuals A and B. For all determiners occurring in natural language, if the determiner relation holds between A and B, it also holds between A and $A \cap B$ (the *live-on* property of Barwise and Cooper (1981)). In DRT, Q expresses a similar relation: the condition is verified by an embedding function f if the generalized quantifier relation for the determiner Q holds between the set of all individuals for which there is a g which extends f to the universe of K_1 and which verifies K_1, and the set of all individuals for which there is an h extending every such g to the universe of K_2 and verifying K_2. Since embedding functions h verify both K_1 and K_2, they are analogous to $A \cap B$. Unlike the singular version of DRT, *every* and *if* introduce different sorts of DRS condition. Conditional sentences produce implicative conditions as before, but *every* is translated into a generalized quantifier $K_1[Every\ x]K_2$. The verification conditions of the two conditions are (contingently) the same. The truth definition makes use of uses Link's theory of plurals (Link, 1983), in which plural collections are represented as members of a lattice structure.

Prior to Kamp and Reyle (1990), it was suggested that DRT would run into the proportion problem described in section 2.3.1 on sentences with quantifying determiners other than *every*, such as

(21) Most farmers who own a donkey beat it.

Richards (1984) shows that the proportion problem occurs if quantifiers such as *most* are treated as *every* was in the singular version, where determiner relations apply to the number of embeddings which verify each sub-DRS, and hence to the number of different assignments to any of the discourse referents. The problem is overcome in plural DRT by using duplex conditions which specify one discourse referent over which the quantification occurs.

Several additional construction rules are added, in particular:

1. Summation, which introduces a new discourse referent equal to the sum of any two or more accessible discourse referents.

2. Abstraction, as explained above, which can apply to any discourse referent of any sub-DRS.

3. Distribution, which allows a condition to apply to each of the atoms making up a plural discourse referent instead of the referent itself.

4. A definite noun phrase rule, which, like the proper noun rule, introduces discourse referents and the associated conditions into the topmost DRS.

One further rule will be discussed below.

Heim's statement of the claims made by DRT remains true in the new fragment. Anaphoric pronouns are still treated as bound variables; indefinites and quantifying determiners are still treated differently, all variables introduced by indefinites being bound either by the quantifying determiner which most closely surrounds them, or given existential quantification by default. An indefinite bound in the former way receives a universal reading, regardless of what the quantifying determiner is. So the reading of

(22) Most farmers who own a donkey beat it.

can be paraphrased as "most farmers who own one or more donkeys beat *all* the donkeys they own."

The distribution rule is optional, and is used with non-quantifying determiners. If this rule is not applied, then the reading for

(23) Three lawyers hired five cleaners.

is that there is a collection of three lawyers who between them hired five cleaners. The distribution rule iterates over the set of three lawyers, giving the reading that each of three lawyers hired five cleaners, possibly different ones for each lawyer. Numerals are normally analysed in the much same way as indefinites, introducing a cardinality condition, but not creating a new sub-DRS. Kamp and Reyle also give an alternative, in which the determiner is analysed as quantifying rather than indefinite. They say that the former DRS is the preferred one. The motivation for this preference appears to come from verbs that do not have a collective reading, which can be conveniently incorporated into the fragment by making the distribution rule obligatory for them (p.760). It follows that the determiners *exactly two/three/...* will never actually be used in their quantifying role – *two/three/...* with the distribution rule will be used instead – and they should perhaps be dropped from the lexicon. After applying the optional distribution rule, a noun phrase with the numeral determiner acts as if it had a quantifying determiner, and any noun phrases occurring within its scope are made inaccessible to anaphora outside. One interesting point about the analysis is that it leaves numeral determiners as indefinite in the sense that there is some indeterminacy about their referent, whilst allowing them to act as if they were quantifying, both in terms of the interpretation they receive, and in terms of the accessibility relation they induce.

The collective reading is the C_1 version, with no provision for C_2. Kamp and Reyle also choose to omit the cumulative reading of (23), according to which a total of three lawyers between them hired a total five cleaners, without them necessarily acting as a group. A brief sketch of how it might be done and the difficulties involved is given (pp.763-765).

The annotation *pl* is required to allow

(24) Three lawyers hired a secretary they liked.

on a distributive reading, where each lawyer hired a secretary he personally liked. The plural pronoun comes under the scope of the distribution, and so refers to an individual. In this example, there will be a plural discourse referent for *three lawyers*, and an individual discourse referent which iterates over its members. The individual discourse referent is marked *pl*, so that *they* can access it. *pl* thus expresses that there must be syntactic number agreement, but not a semantic one.

There is one more rule of the construction algorithm which is of note here. It is needed for dependent plural pronouns, as in:

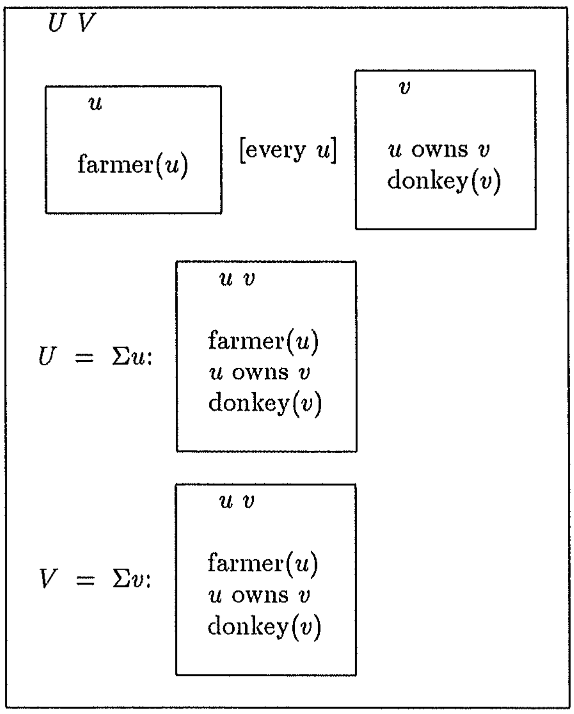
(25) Every director gave a present to a child. They opened them right away.

where the second sentence means that each individual child opened the present given to that child. A further construction rule is needed, called "distribution over abstraction", which allows the abstracted collection of children to be broken down into its members. A change has also to be made to the *pl* annotation to indicate that *them* has a plural antecedent, which can be broken down into individuals as a result of the presence of another antecedent (*they* in this case). The notation used is *pl(u)*; in the case of (25), u would be the antecedent for the individual children derived using distribution over abstraction.

Some sample DRSs formed using Kamp and Reyle's rules are as follows. The DRS for

(26) Every farmer owns a donkey

is



The upper part of the DRS represents the quantification over farmers. The abstracted DRSs in the lower condition ($U = \Sigma u : \ldots$ and $V = \Sigma v : \ldots$) represents the collection of farmers and of donkeys. The sub-DRS in the abstractions are formed by merging the two sub-DRSs of the quantification.

For a collective reading of a plural pronoun, as in

(27) They are cruel.

The DRS becomes:

$U\ V$
...
U are cruel

In this and the following DRSs, "..." represents the quantification condition and the abstractions, which are exactly as in the previous example. For a distributive reading, the abstraction is taken apart using the distribution over abstraction rule. The DRS for

(28) They beat them.

is (slightly simplified):[3]

<table>
<tr><td colspan="3">$U\ V\ W$</td></tr>
<tr><td colspan="3">...</td></tr>
<tr><td colspan="3">$W = U$</td></tr>
<tr>
<td>$w^{pl}\ v^{pl(w)}$

$w \in W$
farmer(w)
w owns v
donkey(v)</td>
<td>[every w]</td>
<td>x

$x = v$
w beat x</td>
</tr>
</table>

The DRS is read as saying for every member w of the plural discourse referent W, the conditions hold. That is, every member of W is a farmer who owns a donkey, and who beats all the donkeys he owns. The conditions on the left hand side of the quantification over w are obtained by copying them from the DRS abstracted over u above.

3.1.5 Discussion

Although DRT has had many successes as a theory of anaphora, a number of criticisms can be levelled at it, on both empirical and methodological grounds.

3.1.5.1 Readings of donkey sentences

DRT assigns the universal reading to quantified donkey sentences, and the symmetric one to conditional donkey sentences. Kadmon (1990) shows how her uniqueness analysis, described in section 2.3, may be incorporated into DRT in a straightforward way. When a uniqueness condition is required, triggered by a definite NP, then an extra condition is entered into the DRS containing the antecedent. For example, in

[3]Formed by applying the rules CR.PRO[Num = plur] to the subject, followed by optional distribution, followed by CR.DA for distribution over abstraction, and finally CR.PRO[Num = plur] to the object. See appendix A for details.

(29) Most women who own a dog talk to it. (Kadmon, 1990, p.311)

an extra sub-DRS is introduced into the left hand side of the duplex condition, with a condition similar to a Russellian definite description. If x is the discourse referent for *woman* and y for *dog owned by* x, then the extra sub-DRS says that if z is a dog owned by x, then $z = y$, for all z.

DRT is perhaps too inflexible in this respect. In the Kamp and Reyle version, only the universal reading is available, and in the Kadmon one, only the unique antecedent reading. As we saw in section 2.3, there has been considerable debate about the correct reading for donkey sentences. Kamp and Reyle themselves acknowledge that there is a problem with

(30) Most farmers who own a donkey beat it.

of which they say

> judgements become unstable and inconsistent as soon as the situation includes farmers who own more than one donkey. (p.771)

The problem could be solved by allowing DRT to produce representations for the different readings, perhaps with some ordering to indicate a preferred one. However, this is a problem with doing so. Suppose one reading entails a contradiction, as the universal reading of "daughter" sentences such as (31) does.

(31) Every man who has a daughter thinks she is the most beautiful girl in the world.

Since the contradiction is a consequence of the specific lexical items *think* and *most* involved and the relation between them, there must be some way of recognising this, so that the DRS can be eliminated from the representation of the whole discourse, and subsequent reference to it prevented. After an attempt to give a universal reading to (31), we do not want to allow

(32) They love their fathers for it.

with the meaning that every daughter loves her father. Operations on the representation form part of the construction process in DRT. Consequently, if the contradiction is spotted in the interpretation, when the DRS is verified, there has to be some way that the verification conditions can "signal" the construction algorithm to discard the DRS. An alternative is to associate DRSs which stand for meaning postulates with lexical items; for (31), something to indicate (ultimately) that it is impossible to think more than one girl is the most beautiful one. However, even if this is done, some logical inference is needed, for example to put the meaning postulate together with the knowledge that men can have more than one daughter. In brief, it may not always be possible to keep D-relations and T-relations distinct while still capturing the range of donkey sentence readings.

3.1.5.2 Problems with plural DRS constructions

There are three empirical problems with plural DRS constructions. Firstly, a distributive reading of

(33) Every farmer owns a donkey. They beat them.

where each farmer beats the donkey he owns is possible, as is a collective one where the collection of farmers as a whole beats the collection of donkeys as a whole. However, it is not possible to obtain a DRS which represents a third case, where each farmer beats some donkey, not necessarily his own, and each donkey is beaten by a farmer, not necessarily its owner. Although this reading is entailed by the collective one, it does convey more information. For example, if the discourse continued with

(34) They hate them.

a possible reading has each farmer hating the specific donkey(s) he beats, rather than the one(s) he owns. DRT as it stands copies all the conditions relating the antecedents into the DRS for (34), so forcing the individuals standing in the *hate* relation to also stand in the *own* and *beat* relations.

A construction rule for this possibility could be added, although it is tricky. We need to copy only the conditions that derive from the nominals and not from the verb that relates them, to apply distribution over abstraction to the plural discourse referents for both the farmers and the donkeys, and to associate a donkey with each farmer and vice-versa. For example, in place of the DRS for (28) shown above, we would have (something like):

<table>
<tr><td colspan="3">

$U\ V\ W\ X$

...

$W = U$
$X = V$

</td></tr>
<tr><td>

w^{pl}

$w \in W$
farmer(w)

</td><td>[every w]</td><td>

x

$x \in X$
w beat x

</td></tr>
<tr><td>

x^{pl}

$x \in X$
donkey(x)

</td><td>[every x]</td><td>

w

$w \in W$
w beat x

</td></tr>
</table>

which means that for each farmer, he beats one or more of the donkeys, and for each donkey there is one or more of the farmers who beats it.

A second problem arises with distribution over abstraction. There is a reading of

(35) Three boys buy five roses.

on which, for example, one boy buys one rose, and the two others collectively buy the remaining four. The sentence

(36) They like them.

can then mean that each boy or collection of boys likes the roses bought by that boy or collection. Distribution over abstraction will not yield this reading, since it iterates over each *individual* boy in the collection, and not over the collections of boys defined by their rose-buying activities. DRT's collective reading again entails the reading we want but is less precise than it.

The third problem is one noted in section 2.4, examples (89)–(90). In Kamp and Reyle's theory, (37) is rejected by means of the *pl* mechanism, which is determined by the syntactic structure of the sentence.

(37) Every director gave a present to a child from the orphanage. One of them opened ?them straight away.

However, it is possible to make the sentence acceptable in a suitable model, for example one where there are more directors than children.

One methodological point also arises. DRT's treatment of plurals requires numerous technical devices to obtain the right readings. To recap, there are operations of summation over discourse referents, abstraction of discourse referents from DRSs, distribution over plurals, distribution over an abstraction, and the marking of individual discourse referents with *pl*, and *pl(u)*. Furthermore the rules as given by Kamp and Reyle have few restrictions on when they may be applied, and so potentially we could form the sum of any two or more discourse referents occurring in the universes of one or more DRSs, the abstraction of any discourse referent from any DRS, and so on. As Kamp and Reyle themselves say:

> The second sentence [(25)] has, if we set plausibility aside, a great many readings, as each of the pronouns *they* and *them* can refer to each of the sets that can be obtained by Abstraction from the duplex condition introduced by the first sentence.

From a technical standpoint, this is not an important criticism: to deal with a complex phenomenon, we may need to employ complex formal devices. A more serious criticism comes when we put the theory together with the claims about DRT as a theory of verbal communication and cognitive representation. If this claim is accepted, then it necessary to accept one of two things. Either cognitive agents carry out some counterpart of the copying and manipulation operations on DRS conditions and discourse referents; or these operations must be considered as purely technical devices, while other DRS operations do have cognitive validity. There is, at present, no evidence to settle the first point either way, although it seems a little unpalatable given the wide range of operations proposed. If the second claim is accepted instead, it raises the question of where the dividing line between the purely formal operations and the cognitively valid ones lies. A way of maintaining DRSs' philosophical status might be to look for a different (non-DRS) semantic mechanism for the anaphoric part, leaving DRSs as a representation purely for the communicative and cognitive parts of the theory.

3.1.5.3 Indefinites as variables

Chierchia and Rooth (1984), in their discussion of configurational notions in DRT, identify the treatment of indefinite noun phrases as being, together with that of conditionals, "the heart of Kamp's proposal". To reiterate the claims of Kamp and Heim, indefinites are not quantifiers in themselves. The quantificational force they receive comes from the quantifiers or conditionals that enclose them, and only if there are none do they have existential force. It is this which is one of the central contrasts of the DRT approach with Montagovian theories of semantics, and the work in anaphora of Evans, Cooper and others.

Empirically, the need to treat indefinites as free variables originates in the universal reading of quantified donkey sentences. In order to obtain this reading, the indefinite *a donkey* must by treated as if it were universally quantified. As we have seen, the empirical evidence does not conclusively support this, and principles like Kadmon's are needed for to obtain other readings. There seem to me to be two objections to indefinites as variables on grounds of simplicity. The first objection is a reductionist one. Indefinites as variables happens to fit (some of) the data for anaphora under quantification, but there is no independent empirical evidence, from other linguistic phenomena, to support it. If there is some other means of explaining the data which allows indefinites to be existentially quantified, then it might be preferable to adopt it. Secondly, indefinites as variables may be hard to integrate with an incremental theory of semantic processing. Theories of this sort are discussed by Haddock (1988); the idea is that both intuition and psycholinguistic studies suggest that the process of understanding an utterance takes place in small, gradual steps as each word is processed. It is hard to reconcile this with indefinites as variables: the indefinite NP in question cannot be processed at all until it is known how it fits into the DRS configuration, and hence what its quantificational force is.[4] The same argument perhaps also undermines Kamp's claim that DRSs are partial models. Certainly, they are partial compared to the models of many logics used in natural language semantics. However, if we cannot exploit this partiality because of not knowing what existential force the discourse referents of the DRS will ultimately receive, its value is weakened.

A problem with indefinites as variables has also been brought to light by an attempt to formalise the partial nature of DRSs. Cooper and Kamp (1991) discuss how the insights of DRT and Situation Semantics might be brought together, an enterprise that has been considered since the inception of the two theories in the early 1980s. Situations, as conceived by Barwise and Perry (1983), represent partial information in a fine-grained way, and so should be ideal for giving a semantics to the partial structures of DRT. The paper presents a simple DRS language, in which the only condition involving sub-DRSs is negation, $\neg K$. The verification conditions resemble those of singular DRT, except that verification is expressed with respect to a situation s rather than a model. This can be understood as either the situation verifying the DRS or as the DRS describing the situation. The initial proposal is that $\neg K$ is verified if there is no situation co-actual with s and no anchoring (embedding), which together verify K. The proposal is unsatisfactory in that it is non-local; that is, the situation under consideration, s, is of no more or less importance than any of the other (co-actual) situations quantified over. This goes against the motivation of using a situational account, where information is kept local and partial as far as possible.

Cooper and Kamp hence look for a more local statement. Three possibilities are

[4] It may be interesting to note that Haddock's theory does have something in common with indefinites as variables, in that the indefinite article relies on implicit existential quantification of the variable it introduces (Haddock, 1988, p.101). The definite article introduces a "closure constraint", which forces its referent to be unique, i.e. for there to be only one possible referent for it when all constraints on the NP have been evaluated. It might be suggested that *every* could function in a similar way to definite articles, but using a different closure constraint. However, the problem is that the closure constraint from *every* applies not only to its own referent, but to the referents of all NP occurring within it, for example in relative clauses, or in the phrase it is applied to, for example the VP. Definite articles apply the closure constraint only to their own referents.

considered.

First approach. Take out the quantification over situations altogether. The result is that negation is no longer persistent, i.e. there can be a situation s which verifies $\neg K$, and a situation s' containing s, which verifies K. This is not in accordance with intuitions about negation.

Second approach. Introduce both positive and negative verification relations. As well as doubling the number of verification conditions, the clause for $\neg K$ in the negative verification relation requires universal quantification over possible embedding functions, which in turn requires that situations contain a large number of negative facts. Cooper and Kamp quote the example of *John doesn't own a car*, which could only be given a negative verification on this approach if the situation being examined contained, for every car, the fact that John did not own it.

Third approach. Introduce quantified facts, or *infons*. The quantification may be of two sorts. We could try quantification over an infon containing a variable: in the above example, $\text{NOT}(\exists x \ll car, x; 1 \gg \wedge \ll own, john, x; 1 \gg)$. But this proves to lead back to same problem as the last solution, requiring facts about many possible values for x. The only possibility that remains is to use a generalized quantifier infon, which has two argument positions for properties. There is independent motivation for infons of this kind in situation theory. For *John doesn't own a car*, we end up with $\ll Exist, P, Q; 0 \gg$, where, roughly, P is the property of being a car, and Q is the property of being owned by John, and the infon says that it is false that there exists an object with both properties. This infon contains an existential quantifier, and we have hence arrived at a quantificational treatment of indefinites.

Summarising, the attempt to give a situation semantics to a DRS language with negation entails one of three things:

1. giving up the persistence of the content of negative sentences.
2. using situations containing a large number of negative facts.
3. giving up "indefinites as variables", and moving to a existential quantification treatment.

Neither of the first two options is desirable, and if we adopt the third we are undermining one of the key ideas of anaphoric DRT.

Neale (1990, chapter 6), as part of a wide-ranging discussion of ideas arising from Russell's theory of descriptions, shows how the universal reading can be obtained without adopting indefinites as variables. He analyses pronouns in donkey sentences as "D-type", which for the present purposes can be thought of as broadly similar to E-type (see section 2.1). D-type pronouns are translated by means of a quantifier *whe* ("whatever") which quantifies over all individuals that satisfy the predicate that follows. For

(38) Every farmer who bought a donkey vaccinates it.

he obtains (p.236)

[every x: man x & [a y: donkey y] (x bought y)]
    [whe z: donkey z & x bought z](x vaccinated z)

which can be paraphrased as "for every farmer who owns one or more donkeys, whatever donkey(s) that farmer bought, he vaccinated the donkey(s)". This leaves Neale with the problem of explaining how the singular *it* can translate into something which acts like a plural. His answer is that D-type pronouns agree with their antecedents syntactically, but need not do so semantically. Instead they are ambiguous between numberless and singular interpretations, the former being realised in the logic by *whe*. The syntactic agreement is necessary to explain the unacceptability of

(39) ?Every farmer who bought a donkey vaccinates them.

where *them* will receive the same semantic translation as D-type *it*. Further evidence comes from

(40) Every farmer who bought more than one donkey vaccinated it.

which has a singular pronoun in agreement with its antecedent, but which has to have a plural reading. The ambiguity between numberless and singular forms of *it* explains why there are different intuitions about the meaning of donkey sentences. If the pronoun is taken as being singular, a unique antecedent reading is obtained; if numberless, a universal reading. For deciding between the readings, Neale suggests that "various contextual and linguistic factors" are used, though he makes no suggestion about what such factors might be. Nor does he explain how to obtain the predicate over which *whe* quantifies – "donkey z & x bought z" in the above example. However, his work provides evidence that the Kamp/Heim commitment to indefinites as variables can be weakened without losing the empirical predictions of DRT.

3.1.6 Conclusions

DRT has provided insights, both technical and philosophical, into a variety of semantic phenomena. The most detailed work has been in anaphora, and the current version of the theory has far wider coverage than any other theory in this area. However, it does face certain problems: it is less flexible than is required to give an adequate account of anaphora under quantification and of certain examples of plural anaphora. More serious is DRT's treatment of indefinites as not carrying any intrinsic quantificational force, which forms a central part of the theory. Besides contributing to the empirical problems with anaphora under quantification, it is hard to reconcile indefinites as variables with the claim of DRSs as partial representations, from the point of view of both incremental interpretation, and of finding a semantics for the representations which is also partial.

If it is possible to find an alternative to indefinites as variables, the major thing that is left of DRT's theory of anaphora is the use of structural relations to define accessibility of antecedents to anaphors. In section 2.4, it was suggested that a semantic, non-structural account of accessibility might be possible. If this is so – a proposal which is tested out by the theory of chapter 4 – then the anaphoric part of DRT would appear to be dispensible.

3.2 Dynamic Logics

In natural language semantics, the familiar view of a logic is as a formal language, the formulae of which are interpreted to supply some informational value, typically truth conditions relative to a model or set of possible worlds. Dynamic logics interpret a formula as a relation between information states, specifying the information present after interpreting the formula, given the information present before its interpretation. This view makes them obvious candidates for a logic of anaphora, with the change in information being identified with the contribution of an antecedent to the discourse. The first such logic was Groenendijk and Stokhof's Dynamic Predicate Logic, which provides a compositional alternative to the original version of DRT described in Kamp (1981). A more recent development, Dynamic Montague Grammar, has attempted to apply dynamic interpretation to sub-sentential expressions, and allows more fine-grained interpretation. The Dynamic Type Theory of Chierchia (1991) follows in a similar vein. The three theories are discussed in the following sections. A final section surveys some related work.

3.2.1 Dynamic Predicate Logic

Dynamic Predicate Logic (DPL), described by Groenendijk and Stokhof (1991b)[5] takes the language of first order logic (FOL) and defines a new interpretation for it, in which existential quantifiers may bind variables that are outside their normal scope. To understand why this is useful as a logic for anaphora, consider

(41) A man walks in the park. He whistles.

The straightforward compositional translations of the first sentence is

$$\exists x[man(x) \wedge walks(x)]$$

For the second sentence, we could translate the anaphor as a free use of the variable x, giving

$$whistle(x)$$

Conjoining the translations, we get

(41a) $\exists x[man(x) \wedge walks(x)] \wedge whistle(x)$

DPL differs from FOL in allowing the variable in the final conjunct to be bound by the existential quantifier, even though it lies outside the quantifier's scope, so that in the example the binding of a value to x established by the existential quantifier still holds in $whistle(x)$. Certain constructs can close off the binding: if the existential quantifier is contained within a universal quantification, negation or implication, then the variable loses its binding outside the construct. Thus, the DPL translation of the donkey sentence (42) is (42a).

(42) Every farmer who owns a donkey beats it.

(42a) $\forall x[farmer(x) \wedge \exists y[donkey(y) \wedge own(x, y)] \rightarrow beat(x, y)]$

The binding for y holds in $beat(x, y)$, but not outside the formula as a whole.

In the semantics of DPL, bindings to variables are maintained by means of assignment functions, i.e. mappings from variables to values. A formula is interpreted as a set of pairs of assignment functions, giving all of the possible "output" bindings for each possible "input" binding. Assignments can be thought of as the information states of the logic. To illustrate how this works, the clauses of the interpretation function for the existential and universal quantifiers are:

[5] Also see Groenendijk and Stokhof (1987) and Groenendijk and Stokhof (1988), which contain earlier versions of the same material.

$$\|\exists x\phi\| = \{\langle g,h\rangle \mid \exists k : k[x]g \;\&\; \langle k,h\rangle \in \|\phi\|\}$$

$$\|\forall x\phi\| = \{\langle g,h\rangle \mid h = g \;\&\; \forall k : k[x]g \Rightarrow \exists m : \langle k,m\rangle \in \|\phi\|\}$$

where $k[x]g$ means that k differs from g at most in its binding for x. The existential clause finds an assignment k which has a binding to x, and such that if ϕ is interpreted with respect to k, the assignment h results. Universal quantification make the final binding the same as the initial one ($h = g$), and requires that every possible assignment derived from g by binding x satisfies ϕ, i.e. some output binding from ϕ can be found. Truth is a derived notion in DPL. A formula is true with respect to an input assignment g if there is some output assignment h such that $\langle g,h\rangle$ is in the interpretation of the formula.

An atomic formula imposes extra conditions on assignments:

$$\|R(t_1,\ldots,t_n)\| = \{\langle g,h\rangle \mid h = g \;\&\; \langle \|t_1\|_h,\ldots,\|t_n\|_h\rangle \in F(R)\}$$

where $F(R)$ is the extension of the predicate R. For conjunction, the "output" assignments of the left hand conjunct are matched with the "input" assignments of the right hand one:

$$\|P \wedge Q\| = \{\langle g,k\rangle \mid \exists h : \langle g,h\rangle \in \|P\| \;\&\; \langle h,k\rangle \in \|Q\|\}$$

Note that this means that conjunction is not commutative. The conjunction rule can be understood as saying that there is an assignment resulting from making P true which also makes Q true. Implication leaves the assignment unchanged, and requires that for every way of satisfying the antecedent of the implication, there is some way of satisfying the consequent:

$$\|P \rightarrow Q\| = \{\langle g,k\rangle \mid k = g \;\&\; \forall h : \langle g,h\rangle \in \|P\| \Rightarrow \exists j : \langle h,j\rangle \in \|Q\|\}$$

As an example, consider (41)

(41) A man walks in the park. He whistles.

which was translated as

(41a) $\exists x[man(x) \wedge walks(x)] \wedge whistle(x)$

The interpretation of the first conjunct is

$$\{\langle g,h\rangle \mid h[x]g \;\&\; h(x) \in F(man) \;\&\; h(x) \in F(walk)\}$$

That is, the set of pairs of assignments $\langle g,h\rangle$ such that h differs from g in its binding for x, with $h(x)$ being a man who walks. In procedural terms, if g is some assignment from preceding sentences in the discourse (an empty assignment at the start), it will be replaced by an assignment differing only from g in binding x to such an individual. The second conjunct has the interpretation

$$\{\langle h,k\rangle \mid k = h \;\&\; k(x) \in F(whistle)\}$$

which leaves the assignment unchanged, but blocks any assignments where the binding for x does not whistle. The interpretation of (41) as a whole is formed by matching up the h assignments in the two sets, following the conjunction rule. After simplifying, the result is:

$$\{\langle g,h\rangle \mid h[x]g \;\&\; h(x) \in F(man) \;\&\; h(x) \in F(walk) \;\&\; h(x) \in F(whistle)\}$$

For donkey sentences, translated as (42a), the interpretation is

$$\{\langle g,g\rangle \mid \forall h : h[x,y]g \;\&\; h(x) \in F(farmer) \;\&\; h(y) \in F(donkey) \\ \&\langle h(x),h(y)\rangle \in F(own) \;\Rightarrow\; \langle h(x),h(y)\rangle \;\in F(beat)\}$$

which is the universal reading of the sentence. In the interpretation of $beat(x,y)$, x is still in scope and its binding can be found just as in FOL. y receives its binding dynamically. Outside the formula, neither binding is visible, since the interpretation clause for $\forall$ ensures that assignments after the formula are the same as those before. Hence *It brays*, translated as $bray(y)$, is uninterpretable, or at least y does not have the correct binding.

DPL has the same coverage as singular anaphoric DRT, and is subject to the similar empirical criticisms. DRT's accessibility relation has a close counterpart in DPL's notion of binding; for example, both can be defined structurally on the syntax of the logical language, or in terms of partial assignments. Groenendijk and Stokhof claim that DPL represents a methodological improvement over DRT in two ways: it is compositional, and it is non-representational. I now look at these claims in detail.

3.2.1.1 Compositionality

To evaluate the first claim, we need to recap what it means for a theory to be compositional: namely that the meaning of a complex expression is derived solely from the meaning of its parts and their method of composition. In natural language semantics, this typically means that when two constituents are combined syntactically, the semantic translation is formed by combining their translations, without drawing on information from any other source. There are two reasons for wanting a compositional theory: to abstract away from how the translation of a constituent was built up from its subconstituents, and to allow us to enumerate everything which contributes to the meaning of a constituent.

DRT is largely in keeping with this principle. It certainly does not draw on information outside the constituent being processed, except in the selection of an appropriate antecedent for an anaphor. The major difference between the theories is that DPL takes a bottom-up approach: it first forms the translations of the lexical items, and then combines them, guided by the structure of the constituents in which they occur. DRT's translation procedure is top-down: it starts from a parse tree of a whole sentence, and progressively rewrites it, generating DRS conditions as it does so. In this respect, DPL can be said to follow a stricter form of compositionality. The translation of any constituent can be given a denotation as soon as it is formed; for example, the denotation of a verb phrase is the set of individuals which allow assignments to be extended through the verb phrase and the assignments that result in each case.[6] DRT cannot guarantee to give any constituent a denotation until the construction algorithm has terminated. For example, after applying DRT's rule for simple *NP VP* sentences, the result is a DRS containing the parse tree for the NP labelled with a new discourse referent, and the original sentence with the NP replaced by the discourse referent, neither of which has a denotation. Similar remarks apply for processing a VP containing a transitive verb. Although the theories will arrive at the same empirical predictions, the DRT construction algorithm may be harder to give an incremental interpretation to than the compositional translation procedure for DPL.[7]

[6] This is not actually how DPL's interpretation is formulated, but it can be expressed in these terms; compare Barwise (1987) or Rooth (1987), discussed below.

[7] For some discussion of compositionality in DRT and DPL, also see Groenendijk and Stokhof (1991b) and Kamp (1990).

3.2.1.2 The nature of representation

Groenendijk and Stokhof's other claim in favour of DPL over DRT is that it is non-representational. To understand this, we first need to make clear what is meant by a "representation", something which Groenendijk and Stokhof are a little elusive on. Two possible definitions are:

- A semantic theory is representational if the semantic representations (formulae, etc) it uses are (abstractions of) psychologically real entities.
- A semantic theory is representational if the semantic representation of one part of a discourse cannot be discarded without making it impossible to interpret the discourse as a whole.

In both cases, the system of representation contains information which is needed to interpret an utterance and which is not present either in the utterance itself or in the underlying knowledge, for example the model. For example, in *he whistles*, the utterance itself does not tell us who *he* is, and the model tells us only what *whistle* means. In section 3.1, we saw that DRT has been treated as representational in both senses. Kamp and Guenthner have proposed that DRSs represent, at an abstract level, elements of thought, and that the way they are constructed and relate to one another mirrors how language is processed. The DRSs used for anaphors need not be considered psychologically real, but they are certainly representational in the second sense. A discourse cannot be interpreted without DRSs to tell us what discourse referents are accessible and what entities they can refer to. The only way in which we can dispense with DRSs and arrive at a non-representational version of DRT is to assume that we can obtain an entire discourse at once, process it to yield truth conditions, and then never refer to anything in the discourse again.

Groenendijk and Stokhof make it clear that they are opposed to semantic theories that require this level of representation. There are three main objections, as presented in Groenendijk and Stokhof (1991b, section 5.2). The first is parsimony. Unless there is a strong reasons for using representational devices, one should not do so, for reasons of keeping the theory formally simple and its insights clear. A second objection is raised against those representational theories which posit some sort of psychological reality to the representation. A weak statement of the psychological view claims that there are such things as mental representations, and they may or may not correspond to representations in a semantic theory. A stronger version claims that not only do psychologically real representations exists, but that the semantic theory works with them, at a greater or lesser level of abstraction, and their existence is necessary for the theory to succeed. The weak statement is of limited scientific value, in that it leaves the semanticist free to choose whatever semantic theory they like. Groenendijk and Stokhof consider the stronger claim to be unwarranted: there is simply not sufficient evidence to establish that it is so. Finally, they consider the argument that a representational level is needed because natural language is not sufficient to convey meaning. That is, the representation clarifies aspects of meaning which cannot be brought out from the linguistic form of the discourse, with contextual factors made explicit. They dismiss this argument by saying that semantics must start from the premiss that

> natural language is all right. If anything is a perfect means to express natural language meaning, natural language is. (p.97)

a remark which seems reasonable, albeit of little help to the semanticist.

I agree with Groenendijk and Stokhof in rejecting a strong theory of representation for lack of evidence, and a weak one for lack of scientific value. Their last point only makes sense when some model of sentence processing is adopted. Clearly, a sentence such as *he whistles* does not contain enough information in its literal, linguistic form for it to be fully interpreted. Regardless of whether it needs a psychologically real representation, some context is necessary to interpret the pronoun. Understanding where this comes from leads to a clearer idea of what a representation, in a non-psychological sense, is.

Consider how a sentence might be "understood" by a model-theoretic semantics, in a simple (and simple-minded) framework of sentence interpretation. The sentence is first translated from a natural language utterance to a formula of the logical language. The formula is then evaluated through interpretation rules of the logic. The result is a collection of truth conditions: constraints about how the model must be in order for the sentence to be true. What we do with those truth conditions depends on what we take the purpose of the utterance of the sentence as being. For example, the truth conditions derived from an assertion might be used to change the information about an entity in the model. Thus, if we are processing the sentence *A man is walking*, and we have decided that the phrase *a man* is the topic of the sentence, we must find an individual which the model specifies as a man, and modify the model (or select between possible models) to indicate that the predicate *walk* is true of that individual. Similarly, a sentence expressing a generalisation, such as *All ducks quack* might create a universally quantified assertion linking two predicates in the model. We can look at the model as providing both a source of information – what entities there are and what we already know about them – and as a repository for new information.

Now examine what happens in DPL. As before we translate from natural language sentences to formulae of the logic. The interpretation of a formula yields a set of pairs of assignment functions. Each pair is accompanied by truth conditions which specify when the pair is a correct interpretation of the sentence. Supposing that the sentence forms part of a discourse, we will take the assignment functions passed on from previous sentences, and for each of them apply the truth conditions in the model as above. At the end of the interpretation, changes have been made to the model, and a new collection of assignment functions produced; or, on an alternative view, some possible models and some of the possible assignment functions for each have been rejected. What is striking about this description is that the use to which assignment functions are even though it could equally be considered an element necessary to assigning meaning. The only way we can dispense with them is to retain the DPL formulae or the original sentences, and then recompute them each time we wish to process more of the discourse.

The question of whether we consider DPL to be representational comes down to whether information contained in the model and assignments is considered to be representational. This seems to me no more than a matter of how we choose to use the term. DPL does not posit that assignment functions are representational in the mentalistic sense that has been claimed for DRSs. However, it seems essential to consider assignment functions as representational in the sense that they capture information which is not "stored" anywhere else. If it seems preferable, we could say that the model, and also the assignment functions, are "persistent information". Groenendijk and Stokhof's view of what a model consists of is this:

> The model contains all the elements that are necessary to assign meaning to all the expressions of the logical language in question. It presents the ingredients of an ontology. (Groenendijk and Stokhof, 1988, p.458)

They are not explicit on what they think the status of assignment functions is, even though they could equally be considered an "elements necessary to assigning meaning". The closest they come is in a more general discussion of context, considered as information state, where they say

> Information about the values of variables plays a local, more or less auxiliary role, whereas information about the world is global: the latter is what we strive for, the former is subsidiary to that end. (Groenendijk and Stokhof, 1988, p.462)

That is, the model has a greater persistence than the assignment functions.

Two further points follow when we compare DPL with DRT. DRSs are considered to be representational, by Kamp at least, and so there is no need to treat embedding functions (cf. assignment functions) as persistent. They can always be recovered from the DRSs. Secondly, a parallel can be constructed with the model of communication suggested by Kamp, where the DRSs of the speaker guide generation, and the hearer reconstructs DRSs from the utterance. If we take DPL as an underlying logic of verbal communication, the persistent information must be reconstructed by a hearer, in a similar way. For example, the two-speaker discourse (43) is just as acceptable as the one-speaker one (44).

(43) Speaker A: A man is walking in the park.
Speaker B: He is whistling.

(44) Speaker A: A man is walking in the park. He is whistling.

Speaker B can make anaphoric reference to speaker A's antecedent, and so somehow the assignment functions must be available to B. The point is that all persistent information, and not just that contained in the model, has to be transmitted from one speaker to another.

3.2.1.3 Coverage of DPL

DPL covers the largely the same data as the 1981 version of DRT, in particular being limited to singular anaphors. Quantified donkey sentences receive a universal reading, and conditional ones a symmetric reading. A uniqueness account, as described by Kadmon, could be added to DPL either by changing existential quantifiers to "definite" ones ($\exists!$) when there is an anaphor to them, or directly, by placing a condition on possible assignment functions.

Groenendijk and Stokhof (1988) discuss an empirical problem that arises with negation, defined in DPL by:

(45) $||\neg\phi|| = \{\langle g, h\rangle \mid g = h \ \& \ \neg\exists k : \langle g, k\rangle \in ||\phi||\}$

The definition blocks anaphora under negation, which is correct in cases like (46), on a sentence negation reading.

(46) A man doesn't walk in the park. ?He whistles.

Anaphora will also be blocked under multiple negation:

(47) It is not true that a man doesn't walk in the park. ?He whistles.

However, there is evidence to suggest that anaphora under double negation is sometimes permitted:

(48) It is not the case that Mary doesn't own a car. It is bright red and parked outside.

DPL still blocks anaphora in this case. That is, the definition of negation is too strong. Anaphora under subordination and disjunction are not handled; DPL can cope with neither of:

(49) Every player takes a pawn. He places it on square one.

(50) Either there's no bathroom here, or it is in a funny place.

A final limitation is that DPL allows only first order quantification, and that the model theory is purely extensional. Groenendijk and Stokhof (1988, pp.479-481) give some brief details of "hyperdynamic predicate logic" (HDPL) which attempts to solve some of these problems. A sentence is associated with a function from assignments to sets of sets of assignments. This allows some logical operators to be given a more fine-grained definition. For example, a negated formula $\neg\phi$ now supplies all the properties, in the form of possible assignments, that ϕ does not have, and so double negation can recover the properties it does have. HDPL is a stepping stone to Groenendijk and Stokhof's more recent logic, Dynamic Montague Grammar, which is described below.

A restricted range of plural phenomena is analysed by van den Berg (1990). The logic, Dynamic Predicate Logic with Plurals (DPLP), is not directly a development of Groenendijk and Stokhof's formulation of DPL, but is inspired by many of the same ideas. Information states take the form of pairs of sets of sets of assignments, and the interpretation of a formula maps such a pair to a modified one. The motivation for such complex information states is as follows: firstly, we need sets of assignments where simple assignments were used before to allow for sentences like[8]

(51) Some boys are taking their dogs for a walk. They want to be home with it before dark.

where the second sentence is read as each boy wanting to be home with his own dog(s) before dark. A separate assignment function is created for each boy and dog, with the assignments mapping variables to individuals. An alternative approach of using sets or Link-style sums to represent plurals, and using assignment functions ranging over sums would not have given sufficient richness to the information, since the connection between each boy and his dog is lost. DPL does also maintain connections between individuals like this, but they do not become apparent with singular data. For example in

(52) A man walks in the park with his wife. She talks to him.

each possible assignment function contains one man and his wife, in effect maintaining the mapping between individuals. The outer level of "set-ness" in DPLP corresponds to indeterminacy, just as it does in DPL. Finally, the two elements of the pair consist of those sets of assignments which verify a formula and those which falsify it. This avoids the double negation problem in DPL mentioned above. Negation (roughly) interchanges the verifying and falsifying sets, thus avoiding the double negation problem described above.

Van den Berg has some criticisms of DPLP. If nothing satisfies the restrictor of an existential quantification, then the falsifier consists of all assignments, which is incorrect. For example, in

[8] Some speakers find this sentence odd, and prefer *them* in place of *it*.

(53) Every man walks in the park. They whistle.

which is interpreted by means of the usual identity that $\forall x[P] = \neg\exists x[\neg P]$, if there are no men, then the translation is verified by any set of things that whistle. Secondly, in

(54) John has some sheep. Harry vaccinates them.

the collection passed on to *them* should be the maximal collection of sheep that John owns, but DPLP does not ensure this maximality. Finally, the strict numerical quantifiers such as *exactly five*, whilst they can be formulated in DPLP, do not fit well with the dynamic style.

Two further criticisms may be added. Firstly, van den Berg's own maximality criticism, exemplified by (54), extends to

(55) Every farmer owns a donkey. They are mangy beasts.

Although the existential quantifier clause ensures maximality of the collection of farmers passed on, it does not ensure that the collection of donkeys is maximal. All it ensures is that at least one donkey per farmer is passed on. Secondly, the reading of the boys and dogs example may be too restrictive. There does not seem to be a way of allowing for a reading where each boy takes home a dog, but not necessarily his own one, even though this does seem to be a possible reading of (51). In chapter 4, I will discuss the issue of interdependent antecedent and anaphors to them in greater detail. All that need be noted here is that things are more complex than van den Berg makes out.[9]

3.2.2 Dynamic Montague Grammar

Dynamic Montague Grammar (DMG, Groenendijk and Stokhof (1991a)) is an attempt to integrate ideas from DPL into a Montague Grammar framework. The result overcomes some of the limitations of DPL. Higher order quantification is possible, and dynamic interpretation is taken below sentence level. The logical language of DMG is dynamic intensional logic, DIL, which is based on work by Janssen and resembles Montague's IL. DIL differs from DPL in distinguishing variables, which play the same role as they do in standard MG, and discourse markers, which carry the dynamic information as in DPL. Information is placed in discourse markers by means of state switchers, expressions of the form $\{\alpha/d\}\beta$. State switchers in effect "store" the value of the expression α in the discourse marker d for the course of the expression β. Montague's $^\wedge$ and $^\vee$ operators are present, interpreted respectively as abstraction over states and evaluation at the present state. In Montague's IL, the states were world-time pairs; in DIL, states represent assignments to discourse markers.

DMG translations of sentences contains a variable standing for a continuation, which is used to pass anaphoric information into subsequent sentences. Where there is an antecedent in the sentence, a state switcher is used to store the referent of the antecedent in a discourse marker of the continuation. For example

(56) A man walks in the park.

is translated into DIL as

[9] A newer version of DPLP has recently been published (van den Berg, 1991), which differs from van den Berg (1990) in certain technical details of the logic, and which makes unique anaphor readings of donkey sentences available by means of an alternative translation.

(56a) $\lambda p \exists x[man(x) \wedge walk(x) \wedge \{x/d\}^{\vee}p]$

The last term takes the binding for x and stores it in the discourse marker d in the continuation p. To interpret (57) following (56), the translation (57a) is composed with (56a), first $^\wedge$-raising it. The result is (58), which is equivalent to (58a).

(57) He whistles.

(57a) $\lambda p[whistle(d) \wedge^{\vee} p]$

(58) $\lambda p \exists x[man(x) \wedge walk(x) \wedge \{x/d\}^{\vee\wedge}[whistle(d) \wedge^{\vee} p]]$

(58a) $\lambda p \exists x[man(x) \wedge walk(x) \wedge whistle(x) \wedge \{x/d\}^{\vee}p]$

The semantics of DIL is similar to that of Montague's IL, other than the omission of modal and temporal operators, and the different definition of state. The principal difference from IL is that DIL expressions are evaluated with respect to a case in addition to the model and assignment function. The case functions in exactly the same way as assignment functions, but is defined for discourse markers rather than variables. Groenendijk and Stokhof's statement of DIL semantics is Montagovian in style, i.e. the interpretation is stated essentially as operations on other DIL formulae. An interpretation directly in terms of model-theoretic entities has been formulated by Beaver (1991).

Two auxiliary operators are defined to simplify the translation of natural language sentences in DIL. Up-arrow, defined by

(59) $\uparrow \phi = \lambda p[\phi \wedge^{\vee} p]$

gives either the empty set or the set of all propositions which are true in the state in which ϕ is evaluated. Down-arrow, for which the definition is

(60) $\downarrow \phi = \phi(^\wedge true)$

results in the truth conditions of ϕ with its dynamic effects cancelled; $\downarrow \phi$ is true in a state if ϕ can be successfully processed in that state. $^\wedge true$ denotes the set of all states. The up-arrow and down-arrow operators are used to define discourse equivalents of negation, conjunction, existential and universal quantification, and so on in terms of their static counterparts. For example, negation may be defined as

(61) $\sim \phi = \uparrow \neg \downarrow \phi$

which closes off bindings using $\downarrow$, negates the result, and raises the formula back to a set of propositions with $\uparrow$. The discourse versions of existential and universal quantification, which quantify over discourse markers, are:

(62) $\mathcal{E}d\phi = \lambda p \exists x\{x/d\}(\phi(p))$

(63) $\mathcal{A}d\phi = \sim \mathcal{E}d \sim \phi$

Groenendijk and Stokhof give a grammar for a small fragment of English in Montagovian style, showing how to translate it into expressions of DIL. The indefinite article *a* is translated into $\mathcal{E}$, and *every* into $\mathcal{A}$. The resulting fragment has the same coverage and empirical import as DPL. Groenendijk and Stokhof say that extensions to include quantification and anaphora to object of higher types than just individuals are under way. They also give some tentative illustrations of the benefits of DMG over DPL or DRT. For example, versions of negation and universal quantifications can be defined which are sufficiently dynamic to cope with the double negation example (48) and the subordination example (49):

(64) $\sim \phi = \lambda p \neg(\phi(p))$

(65) $\mathcal{A}d\phi = \forall x\{x/d\}(\phi(p))$

Static negation and universal quantification can be derived from these new forms by closing them with $\uparrow\downarrow$. Dynamic implication and disjunction can be defined similarly. All that remains is to specify when the dynamic versions should be used and when the static ones. Groenendijk and Stokhof suggest that some constructions may be accompanied by monotonicity constraints: the resulting translation must act as an upward monotone quantifier over states. This would, for example, block formulae $\sim \phi$ where ϕ is upward monotone and $\sim$ is interpreted as dynamic negation. To obtain upward monotonicity, the static closure $\uparrow\downarrow\sim \phi$ is needed. Although this may work, Groenendijk and Stokhof express some reservations about whether it is exactly right. Some refinements to their approach, which improve on issues of negation and modal subordination, may be found in Dekker (1991).

The value of DMG is in its similarity with other systems of logic. Concepts such as monotonicity are well understood, and some of their implications for natural language semantics have been explored. In a sense, DMG forms part of a semantic "toolkit", a collection of resources which can be applied in solving semantic problems. This is in contrast with approaches like DRT, where the representation and logic are significantly different from other approaches, and a new body of knowledge about the formalism has to be developed.

One final comment about DMG is that the discussion of representation in DPL applies here also. Cases contain persistent information just as assignment functions did in DPL. The principal change is that there are explicit operations whose semantics manipulates cases, such as state switchers for passing information into them, and $\downarrow$ which resets them.

3.2.3 Dynamic Type Theory

Chierchia's Dynamic Type Theory (DTT), described in Chierchia (1991), is closely related to Groenendijk and Stokhof's DMG. On a technical level, the principal difference is that there are no state switchers, and instead the variables over which quantification and λ-abstraction occur are marked as being either ordinary variables or discourse markers. Variables of the former sort are bound through assignment functions, and of the latter sort through cases. Chierchia defines up-arrow and down arrow-functions in the same way as Groenendijk and Stokhof, and similarly discusses possible definitions of negation, quantifiers and so on. However, where Groenendijk and Stokhof are mostly concerned with logical properties of their system, Chierchia's discussion centres around semantic issues. He has four main topics, two concerned with incorporating the insights of other work in semantics into a dynamic framework, and two with empirical issues.

The first topic is relatively minor. An adequate account of noun phrases requires a generalized quantifier framework, rather than the fixed quantifiers of DPL. Roughly speaking, noun phrase denotations take the form of a set of sets of individuals, representing the set of properties of the NP's referent. There are many advantages to such a treatment, including good accounts of complex determiners and of NP conjunction; see Barwise and Cooper (1981) for an extensive discussion. DTT (and DMG) show that it is possible to represent noun phrases as generalized quantifiers in a dynamic logic.

Secondly, Chierchia discusses adverbs of quantification. As we saw in section 2.3.2, adverbs of quantification behave in some respects like determiners, but quantifying over situations rather than individuals. In DTT, the situations can be identified with cases. Thus

(66) When a man is in the bath, he always sings.

will be analysed as saying that every case satisfying the left hand part of the sentence also satisfies the right hand part of it. In DTT this is expressed as

(67) $every(\lambda c\exists x[man(x) \wedge inbath(x) \wedge^{\vee} c])(\lambda c\exists x[man(x) \wedge inbath(x) \wedge sing(x) \wedge^{\vee} c])$

where *every* is defined as $\lambda P\lambda Q\forall c[P(c) \rightarrow Q(c)]$, and the λ−abstractions are over cases. On this definition all that distinguishes one case from another is the bindings it makes to discourse referents, i.e. the information passed on for subsequent anaphors. The definition gives the symmetric reading, in that we quantify over c, which has a binding for each antecedent. To obtain an asymmetric reading, quantifications of the form $D(A)(B)$ are replaced by ones such as $D(\lambda u[_n A])(\lambda u[_n B])$, where $\lambda u[_n A]$ is defined as $\lambda u[A; \uparrow u = x_n]$. The semicolon operator is dynamic conjunction (sequencing). In effect, the abstraction evaluates A, resulting in a binding for x_n, and then stores the value it comes up with for x_n in the discourse marker u, which is used as the distinguishing element between the sets over which D applies. Chierchia points out that his account of adverbs of quantification is far from a full one, and in particular it ignores their modal properties. He does not give a procedure for deciding which variable to quantify over in asymmetric readings, although he does point to topic-comment structure as possibly providing some answers, and gives some examples to support the view.

The remaining two topics are related: the proportion problem and alternative readings for donkey sentences. To avoid the proportion problem, the abstraction mechanism as described in the previous paragraph is used, which singles out one variable. For donkey sentences, Chierchia proposes that there should be two variants for each determiner: a strong determiner, D^0, and a weak one, D^+. The definitions are:

(68) $D^0(A)(B) = D(\lambda u \downarrow A(u))(\lambda u[every'(A(u))(B(u))])$

(69) $D^+(A)(B) = D(\lambda u \downarrow A(u))(\lambda u \downarrow [A(u); B(u)])$

Here D is the standard static determiner familiar from generalized quantifier theory. *every'* is the adverbial form of quantification used above, i.e. a quantification over cases. (68) is derived from the paraphrase used in strong (universal) readings, for example that *most farmers who own a donkey beat it* is paraphrased as *for most farmers who own a donkey, for every donkey they own, they beat it.* Chierchia's weak reading is the indefinite lazy one in the terms of section 2.3.1. A paraphrase is *the number of farmers that own and beat a donkey is greater than the number of men that own but don't beat one.*

Chierchia discusses at length whether making determiners ambiguous is a good approach, and comes to the conclusion that it is not. He states a dynamic equivalent of the conservativity property for static determiners, $DAB = DA(A \cap B)$, claimed to be a linguistic universal. Of the two definitions above, only D^+ is dynamically conservative. Chierchia therefore treats all donkey sentences as being analysed in terms of the weak determiner, requiring some means of producing strong readings. He proposes that the way to do it is to adopt an E-type account. For the weak reading, "donkey" pronouns are translated as variables bound by the dynamic quantifier. For the strong reading, they are translated into a function applied to other variables, still using the weak determiner, the function yielding all of the possible values for the pronoun under the bindings to variables. On this account, the translation of the standard quantified donkey sentence, reduced to its static form, would be

(70) $\forall x[[farmer(x) \wedge \exists y[donkey(y) \wedge own(x, y)]] \rightarrow beat(x, f(x))]$

Two possibilities for f are suggested: it might map x to a group of objects, namely the donkeys owned by x, or there might be several possible values for f, each binding x to one of the donkeys owned, but otherwise indistinguishable. The latter approach is inspired by ideas of Kadmon – it is essentially the same idea that she used when explaining sage-plant sentences.

A number of criticisms of this approach may be raised. Firstly, in trying to avoid an ambiguity in the determiner, Chierchia seems to have introduced a different ambiguity: between the bound variable and the E-type reading. We could in fact dispense with the former, and incorporate the ambiguity into the choice of f. Using an "$f(x)$" approach throughout also admits the possibility of the unique anaphor reading. The problem remains of how to choose a suitable f, and in particular how to derive it either from the original sentence or from its DTT translation. In this respect, we are just as badly off as we were with the E-type approach of Cooper (1979). As with DPL and DMG, the question of whether cases are representational can be raised; they again contain persistent information. If the E-type account is adopted, we need also question the status of the functions f: are they persistent and derived when the antecedent is interpreted, or are they part of the pronoun's translation? If the former is true, the question of whether they form part of the case arises; if the latter, there is a need for a principle to predict why different possible functions are available in some contexts but not in others.

3.2.4 Related theories

There are a number of related theories which follow some of the ideas presented above. Not all of them are dynamic logics, in the sense of mapping between information states, but all of them employ some notion similar to assignment functions or cases. The coverage and predictions are generally the same as the singular version of DRT.

Barwise (1987) discusses an approach to anaphora drawing on ideas from situation semantics. The interpretation is expressed in terms of a relation between assignment functions, with each clause of the interpretation being specified for both verifying and falsifying cases. This is in keeping with one of the tenets of situation semantics, that information is partial. It is not sufficient just to state truth conditions, and assume the falsity conditions are what is left: both should be stated explicitly. The dynamic interpretation is defined on sub-sentential expressions, as well as formulae. The fragment treated includes dependent NPs, such as *another* or *every other*. *Other* is effectively an anaphoric element, in that it constrains the extension of the NP to be different from one already mentioned.

Rooth (1987) draws on Barwise's formalism to develop a compositional alternative to DRT and Heim's File Change Semantics. As with Barwise's theory, interpretation is taken to below the level of formulae. Each logical expression is interpreted as a *parameterised set*, the members of which are tuples containing the input and output assignments and possibly other information. For example, the interpretation of a NP of the form $a\ \alpha_N$ is

(71) $\{\langle g, x, k\rangle \mid \exists h[\langle g, x, h\rangle \in \|\alpha\| \wedge k = h^n_x]\}$

The last term means that we store the value of x that we get from the interpretation of the noun α in element n of the assignment function. There is a lot in common between Rooth's approach and DPL. Both ultimately come down to interpretation in terms of sets of pairs of assignment functions. Rooth presents three interesting ideas based on his theory. Firstly, he shows that a homomorphism can be constructed from his logic to a standard Montague Grammar interpretation of sentences. The theories differ only on anaphors, which are not fully covered by MG. This gives some reasonable certainty that the approach is as correct as existing theories of sentence interpretation. Secondly, Rooth suggests a means of overcoming the proportion problem, by using equivalence classes in the parameterised sets. Determiners are interpreted as sets of the form $\{\langle g, P, Q, h\rangle \mid C(P, Q)\}$, where P and Q are sets of individuals and C is some condition relating them. To avoid the proportion problem, P and Q are replaced by some appropriate equivalence classes. For example, in

(72) Most farmers who own a donkey beat it.

the equivalence condition would be having the same farmer. A final application is to partitives, such as

(73) Each of the students spoke.

which Rooth analyses in his framework by taking *the students* as denoting a Link-style sum of all the students, which is then broken down into individuals by *each of.* The same approach could be applied to the other theories that have been described here.

The model theoretic reformulation of DIL by Beaver (1991) is stated in terms very similar to Rooth's parameterised sets, and Beaver shows that the two are equivalent. There are differences in the coverage of the DMG and Rooth's work, however, arising from the different translation procedures from natural language to the logical language. DMG makes finer distinctions in meaning by, for example, translating determiners into dynamic quantifiers.

Muskens (1991) draws on ideas from semantics of programming languages. He treats formulae as mapping from one state to another, where a state takes the form of an index. The values of dynamic variables are recorded in stores, which are functions from states to values. Something like the "indefinites as variables" approach of DRT is used, in that there is no explicit quantification over stores. For example,

(74) A farmer owns a donkey.

is interpreted as

(74a) $\lambda ij(i[v_1, v_2]j \wedge \mathit{farmer}(v_1 j) \wedge \mathit{donkey}(v_2 j) \wedge \mathit{own}(v_1 j, v_2 j))$

meaning that the output state j differs from the input state i in stores v_1 and v_2, and that these stores in state j denote the farmer and the donkey. Where the anaphoric information is closed off, as in

(75) Every farmer who owns a donkey beats it.

all store references get eliminated and the state remains the same:

(76) $\lambda ij(i = j \wedge \forall xy((\mathit{farmer}(x) \wedge \mathit{donkey}(y) \wedge \mathit{own}(x, y)) \rightarrow \mathit{beat}(x, y)))$

The coverage is the same as singular DRT.

Finally, some ideas have been proposed by Cooper (1991) for a treatment of donkey sentences within a situation semantics framework.[10] He shows how a number of different forms of each quantifier can be defined in situation semantics to capture different sorts of quantification: over specific properties and generic ones in various combinations. The analysis of donkey sentences quantifies over individuals in two situations: ones where a farmer owns a donkey and ones where a farmer owns and beats a donkey. To establish the anaphoric link, each situation of the latter sort is made a part of a situation of the former sort. Pronouns are analysed as definite descriptions. By keeping the situations small enough, the only thing they can refer to is the antecedent's referent. Weak and strong readings may be obtained by appropriately structuring the collections of situations over which the quantification happens. To obtain the strong reading, there are situations specifying all the donkeys each farmer owns, but no smaller situations specifying only some or one of them. Cooper's account is very much a first draft, but it seems promising. The major problem appears to be in deciding which situations to quantify over: in order to decide, we need to know which reading of the sentence we want in advance of forming the reading.

3.2.5 Conclusions

The goal of the theories described in this section was to take the advances made by Kamp's DRT back into a more conventional framework, either first order logic or Montague Grammar. In doing so, compositionality is guaranteed, and it becomes possible to integrate the theory of anaphora with MG and the insights arising from it, such as generalized quantifier theory. Chierchia's work also suggests that there may be reasons for returning to the E-type treatment of pronouns which DRT and File Change Semantics dispensed with.

Part of Kamp's motivation for DRT was that a significant change in how we represent meaning was needed, from truth conditions to a representation of the discourse. Dynamic logics also change the notion of meaning, but in a more restricted way, to operations on information states. Whereas DRT retains a great deal of semantic information about the discourse, dynamic logics work only with what appears to be essential: for the anaphora, bindings to discourse markers.

The simplicity of information states may prove to be a problem. (126)–(129) of section (2.5.1), for example

(129) A sculptor and her spouse walked into a pub. The man was wearing a grey overcoat.

show that we may need information which is more specific than just what individuals are present in the discourse. For this, the comprehensive representation of DRT seems better. Beaver (1991) also suggest that contradictory discourses might also motivate a richer representation. DMG will simply interpret them as an empty set of propositions, but there may be information that can be recovered and used in the remainder of the discourse. A way of doing this in DMG as it stands is to have information states which are made up of DIL formulae, but we have then arrived at a theory which is as representational as Kamp's. Even if we decide to ignore these problems, it is also unclear that the information made

[10] Some work on other sorts of anaphora in situation semantics, particularly with reference to VP deletion (ellipsis), may be found in Gawron and Peters (1990).

available in DPL, DMG and DTT is rich enough. In the next chapter, I will introduce some examples which show that complex interrelations between antecedents need to be recorded, for which assignment functions or cases are not sufficient. Such problems only become fully apparent when plural anaphora is considered, something which dynamic logics have only touched on at present.

3.3 Summary

In this chapter we have looked at two theories (or groups of theories), which are unconventional compared to the majority of work in formal semantics of natural language. The unconventionality manifests itselfs in a variety of ways, of which two are particularly notable. Firstly, meaning is identified with a change in information rather than with truth conditions. Secondly, the logics employed are quite different from the formalisms of previous semantic theories, relying on structured representational devices in DRT, and on dynamic interpretation in the dynamic logics. The motivation behind the unconventionality is the lack of an adequate treatment of anaphora in the logical frameworks that existed prior to DRT and DPL.

As we have seen, some questions have been raised about the empirical adequacy of the theories, compared to the intuitive consideration of the data that appeared in chapter 2. Certain methodological issues also remain open, such as the indefinites as variables claim of DRT discussed in section 3.1.5.3. Finally, the computational properties of neither theory have been adequately explored. Although there have been some implementations of DRT, they are largely concerned with constructing DRSs during or following the parsing process. The difficulties that might be involved in theorem proving or efficient computational implementation of DRT and dynamic logics have so far received little or no attention.

In the next chapter, I will develop a new theory, which is non-representational, unlike DRT, and employs a static logic, unlike DPL/DMG/DTT. In doing so, I will develop a conceptual model of anaphoric processing which can be formalised in a variety of ways, one leading to my approach, another to DRT-like theories, and a third to theories resembling dynamic logics. The fact that this can be done suggests that no single one of DRT, dynamic logics and my own theory is "right" in an absolute sense, and the choice of which approach to anaphora is adopted in any given circumstances becomes a matter of which fits best with other theoretical and practical considerations.

4 TAI: A model for anaphora and its logic

4.1 Introduction

Chapter 2 was principally concerned with the question of what readings are possible for anaphors, particularly where quantifiers are involved, and chapter 3 with theories which attempts to capture such readings formally. In this chapter, I develop a new theory of NP anaphora. The approach taken is to first propose a general description of what information antecedents and anaphors refer to, using empirical evidence to specify the information in detail without reference to a specific logical formalism. The resulting conceptual model is called the discourse set (DS) model. A logic, L(GQA), which allows this information to be extracted compositionally is then defined. The translation procedure from natural language sentences to L(GQA) formulae follows, and variants of the procedure which make different predictions about sentence readings explored. L(GQA) also provides a testing ground for the issue discussed in section 2.4: whether the antecedent-anaphor relation must be constrained structurally, or whether a semantic account is possible. Finally, special cases of some determiners and the problems that the standard logical connectives raise when anaphora is involved are examined. As a whole, the theory is referred to as TAI: theory of anaphoric information.

4.2 The conceptual model

4.2.1 Representing the context

The conceptual model assumes the existence of a *context*, which contains all the information needed to interpret anaphors, and which is derived from the discourse. Every NP, antecedent or anaphor, has a *slot* in the context, labelled with an *index*. For one NP to be anaphoric to another, the information in their respective slots must be identical, or alternatively they must use the same slot. A theory of anaphora consists of specifying precisely what information occupies the slots in the context, and any constraints which relate the content of one slot to that of another. To formalise the theory, what is needed is a concrete representation of the context and a procedure for relating the context to the discourse.

The context can be thought of dynamically, in which case an antecedent is treated as effectively placing information into a slot, and the anaphor as examining this information, and possibly putting back altered information. A static interpretation is also possible, in which all possible contexts are available at the start of the discourse, and those which are inconsistent with information from antecedents and anaphors are discarded as the discourse is processed.

To give an example, consider a discourse which starts with the sentence

(1) A^i man walks in the park.

The notation used here is that the index of a NP is shown as a superscript on determiners and pronouns, with anaphors having the same index as their antecedents.[1] Before processing (1), the information about slot i is completely unconstrained; after processing it, the information in slot i must describe one object, which is a man, and which walks in the park. Continuing with

[1]A slightly different notation is used in the next section. Indices are omitted from the examples where there is no need to show them.

(1a) He^i is happy.

adds a further constraint, that the individual in question is happy. There are no contexts which can satisfy both these constraints and the ones imposed by (1b–d), either because of logical inconsistency, in the case of (1b), or because of gender or number mismatch between antecedent and anaphor, in the other two cases.

(1b) He^i is unhappy.

(1c) She^i is wearing a hat.

(1d) They^i are wearing hats.

Consequently, there are no contexts which can be derived from a discourse containing any of these sentences following the earlier two, and an attempt to interpret such a discourse will fail.

It may be possible to construct more than one context consistent with a given sentence, representing indeterminacy. For example, on hearing (1), there may be several men who the sentence could refer to. Again there are two strategies, one of making an arbitrary choice and backtracking to change it if it proves to be inconsistent with later information, and one of retaining all contexts which remain possible at each stage.

4.2.2 Singulars and plurals

The next step is to identify precisely what information about each NP must be recorded in the context. When we are dealing only with singular anaphors, the answer seems straightforward: just the individual referred to by antecedent and anaphor. This is certainly sufficient for examples like

(2) A^i farmer owns a^j donkey. He^i beats it^j.

Now examine what happens when we have NPs that denote more than one individual:

(3) Every^i farmer owns a^j donkey. They^i beat them^j.

They in the second sentence refers to all farmers who own a donkey, and *them* to the donkeys owned by those farmers. The most prominent reading of the second sentence (though not the only one) is that each farmer beats the donkey or donkeys he owns. A possible representation is to make slot i contain individual farmers and slot j individual donkeys, and to record the links between the corresponding members. The links must then be respected in interpreting the second sentence. In an informal notation, with $\leftrightarrow$ indicating entities that are linked, and the whole context enclosed in $\{\ldots\}$, the result looks like this (on some arbitrary model):

(4)
$$\left\{ \begin{array}{ccc} i & & j \\ f_1 & \leftrightarrow & d_1 \\ f_2 & \leftrightarrow & d_2 \\ \ldots & & \end{array} \right\}$$

Following (3), we could have (5) and (6):

(5) They^i are cruel.

(6) They^j bray.

which are interpreted by taking the whole of the specified slot as the argument to "are cruel" or "bray", i.e. $f_1, f_2, \ldots$ in (5) and $d_1, d_2, \ldots$ in (6).

The same mechanism will work for quantified donkey sentences such as

(7) Every^i farmer who owns a^j donkey beats it^j.

The context has the same form as above. When interpreting *it*, we will be considering one farmer at a time, and so can retrieve the corresponding donkey(s). For example, the entry in slot j corresponding to farmer f_1 is d_1, and so we test whether f_1 beats d_1. Next f_2 is tested to see whether it beats b_2, and so on.

An alternative to this approach might be to construct a separate "sub-context" for each farmer, and pass on a set of sub-contexts rather than a single one. In the notation used before, the result looks something like this, with each $\{\ldots\}$ enclosing a single context:

$$(8) \quad \begin{array}{lcccl} & i & & j & \\ \{ & f_1 & \leftrightarrow & d_1 & \} \\ \{ & f_2 & \leftrightarrow & d_2 & \} \\ \ldots & & & & \end{array}$$

In the examples we have seen so far, the single context approach is a little more convenient to work with, particularly in cases like (5) and (6), where complete collections of individuals across the separate sub-contexts are needed. However, an approach based on multiple contexts may be more appropriate for conditional sentences. In

(9) If a^i farmer owns a^j donkey, he^i beats it^j.

the antecedent sentence establishes a context in which slot i contains one farmer and j his donkey, as in (3). The consequent is interpreted with respect to the same context. The effect of the conditional is to make the iteration over such contexts explicit.

Anaphors to collectively read antecedents raise further complications. Consider:

(10) Three^i boys bought five^j roses.

(11) They^i took them^j home.

Sentences (10)–(11) have several possible interpretations. The one that is of interest here is the cumulative reading of (10), on which a total of three boys bought a total of five roses. The most likely reading of (11) is that each boy or collection of boys takes home exactly the roses bought by that boy or boys. This tells us more about what the context must contain. There is one slot for the boys and one for the roses. In the earlier examples such as (4), slots were divided up into individuals. For (10)–(11), the division is into entities which may be singular or plural. Each entity in one slot is linked to an entity in the other slot, just as individuals were linked before. To interpret (11), we iterate over each entity in the boy slot i in turn, find the corresponding entity in the rose slot, and test whether the taking home relation holds between them. For example, suppose that boy 1 bought roses 1 and 2, and boys 2 and 3 jointly bought roses 3, 4 and 5. Then the context from (10) is (12), in the notation above:

$$(12) \quad \left\{ \begin{array}{ccc} i & & j \\ b_1 & \leftrightarrow & r_1, r_2 \\ b_2, b_3 & \leftrightarrow & r_3, r_4, r_5 \end{array} \right\}$$

(11) is true if boy 1 took home the roses he is linked with, namely roses 1 and 2, and similarly if boys 2 and 3 together took home roses 3, 4 and 5. The linking of one entity to another is termed *dependence*, and the overall relation between the slots is called the *dependence relation.* The division of a slot into entities defined by the dependence relation will be referred to as partitioning, although it should be noted that the term is being used informally, rather than with its set-theoretic meaning.[2] The partitioning of a slot only matters where there are two or more anaphors, and the anaphors have antecedents that stand in a dependence relation. For example, consider (13) following (10) with the context shown in (12).

(13) They^i went to school.

The reading where the b_1 went to school alone and b_2 and b_3 went to school together is no more prominent than any other, even though the context has slot i partitioned this way.

A phenomenon that supports the notions of dependence and expressing indeterminacy through alternative contexts is "cross-sentential counting". Suppose following (10), we continued with

(14a) Joe liked his one.

(14b) Fred ate his one.

(14c) Bill threw his ones/?one away.

Then, (14c) with *one* rather than *ones* is unacceptable. On the model developed here this is because after (10), we can form any context with three boys and five roses partitioned so that they stand in the relation of buying. After each of (14a) and (14b), some of the possible contexts have to be discarded, namely ones which have Joe or Fred related to more than one rose, assuming that *one* is interpreted as something like finding the singular part of the slot corresponding to the specified member of another. When we come to interpret (14c), the only contexts remaining have Joe and Fred each linked to one rose. Bill must be therefore linked to all of the remaining roses, of which there are three. This is not consistent with the singular use of *one* in (14c). Note that the evidence of (14) also supports the idea of semantic number agreement introduced in section 2.4: the acceptability of *ones* in (14c) is contingent on the context established by the preceding sentences, rather than any structural relation between them.

Anaphors can establish dependence relations between two antecedents that were previously unrelated. (15) is an example:

(15) Three^i boys went to the flower shop. There were five^j roses in the window.

(15a) They^i bought them^j.

(15b) They^i took them^j home.

After (15), there is no dependence relation between slots i and j (or, to look at it differently, any dependence relation between them is possible). After (15a), we require that slot i is partitioned into collections of boys which are linked to collections of roses they bought in slot j. (15b) then means that each boy or boys took home the rose or roses they bought, maintaining the dependence established by (15a). Dependence relations which are established "indirectly" in this way can become arbitrarily complex, as in (16).

[2]See sections 2.6.1 and 4.3.2.8 for some related discussion.

(16) a. Some^i men each own a^j donkey.
b. Some^k women each own a horse.
c. The^i men are married to the^k women.
d. The^k women gave the^i men the^j donkeys as wedding presents.
e. They^k bought them^j from cruel farmers.

The first sentence links slots i and j. In sentence (16c), a relation between each partition of slot i and a partition of slot k is established, and hence indirectly between slots j and k, and these relations are respected in interpreting the final two sentences, which could be paraphrased as

(16d′) Each woman gave the man she is married to the donkey he owns as a wedding present.

(16e′) Each woman bought the donkey which the man she is married to owns from a cruel farmer.

For ditransitive verbs, links must be made between three slots. In

(17) Three^i boys gave four^j girls five^k roses. They^j put them^k in vases.

the second sentence may be read as saying that each group of girls given some roses put their roses in a vase. That is, the dependence between slots j and k is maintained, even though there is no reference to slot i.

In (15) and (16), the readings where the anaphors follow the dependence relations of their antecedents are the most prominent ones. It is also possible to read them in a way where the dependence relations are not respected. In such cases, sentences containing anaphors may introduce a different dependence relation on existing anaphoric information. An alternative reading of (10)–(11) to the one considered before is that the roses taken home by each boy or collection of boys was different from the ones bought by the boy or boys. Subsequent reference, for example by

(18) They^i tend them^j carefully.

could pick up on either the original dependence relation from (10) or the new one from (11). That is, each rose, or collection of them, might be tended by the collection of the boys that bought it, or the collection that took it home. This suggests a change to the indexing model suggested above. Anaphors now need two indices, one functioning as before – an "output" index – and one indicating the source of the information – an "input" index. Although the input and output slots have the same content, their partitioning maybe different. Partial full NPs may also remove some of the content of the slot:

(19) Some animals live on a farm. The sheep are plotting to escape. They have already dug a tunnel.

The sheep is anaphoric with antecedent *the animals*, and places the members of the animals slot which are also sheep into the context for *they* to pick up.

One antecedent can stand in different dependence relations to two or more others, as in (20) with both *buy* and *own* read cumulatively.

(20) Three^i boys who buy five^j roses own four^k vases.

Suppose that Joe bought one rose, and Bill and Fred two each, and that Fred and Joe each own one vase, and Bill owns two. Then,

(21) They^i put the^j roses in the^k vases.

can mean that each boy put the roses he bought into the vases he owns. The context in this case is:

$$(22)\quad \left\{\begin{array}{ccccc} i & & j & & k \\ j & \leftrightarrow & r_1 & \leftrightarrow & v_1 \\ f & \leftrightarrow & r_2, r_3 & \leftrightarrow & v_2 \\ b & \leftrightarrow & r_4, r_5 & \leftrightarrow & v_3, v_4 \end{array}\right\}$$

Now consider a different situation, where Joe bought one rose, and Bill and Fred jointly bought the other four, and that Fred owns one vase, and Joe and Bill jointly own the other three. Diagrammatically:

$$(23)\quad \left\{\begin{array}{ccccc} i & & j & & k \\ j & \leftrightarrow & r_1 & & \\ b, f & \leftrightarrow & r_1, r_2, r_3, r_4 & & \\ f & & & \leftrightarrow & v_1 \\ j, b & & & \leftrightarrow & v_2, v_3, v_4 \end{array}\right\}$$

A reading of (21) where each partition of the boys takes the roses owned by that partition and puts them in the vases owned by that partition – possibly paraphrased as *the boys put their roses in their vases* – is not possible, since the partitions do not match up. Note that other readings are still possible.

Finally, the context may contain slots that are empty, as from

(24) No^i man owns^j a donkey.

Slot i contains all the men who own a donkey, i.e. none at all, and slot j all of the donkeys so owned, again being empty, so that no reading is possible for any of

(24a) They^i beat them^j.

(24b) They^i are cruel

(24c) They^j neigh

The same thing happens with the sentential negation reading of

(25) It is not true that a man owns a donkey.

A final related case is

(26) Some^i men own no^j donkeys.

where slot i would not be empty, but slot j would. In this case, the partitioning of slot i is arbitrary: since there is nothing in slot j for any partition of slot i to be linked to, there is no dependence relation between them. Following (26), (24b) is possible, but neither (24a) nor (24c) is.[3]

[3] An alternative behaviour for determiners such as *no* is considered below.

4.2.3 Indexing

The approach to indexing assumed in most of the examples has been that anaphors and antecedents share an index. As already mentioned, we can also assign a different index to every NP, and make anaphoric links by requiring that the input slot of an anaphor is identical in content and partitioning to that of its antecedent. There is an advantage to this approach, in that the anaphor *resolution*, the process of choosing an antecedent, can be captured in the same framework as the detailed analysis of what anaphors mean. Where there is more than one possible antecedent, the contexts for each possibility are set up. Contexts with the wrong choice may be discarded on the basis of subsequent sentences, just as a range of contexts may be set up and then reduced when there is some indeterminacy about the referent of an antecedent.

4.2.4 Extracting the information

Having identified what information the context must contain, we next need to examine where the information comes from. The context is defined by a number of constraints derived from the sentences, which come in two sorts: *content constraints* specifying the membership of a slot, and *structure constraints* specifying partitioning and relations between slots. To illustrate: in

(27) Every^i farmer owns one^j donkey.

the content information is that slot i contains all the farmers who own a donkey, and slot j all donkeys owned by them. The structuring information links each farmer to the donkey he owns.

The next two sections examine the principles which define the content and structure information in general. For the content, this comes down to establishing the relation between the denotation of a given NP and what the slot should contain. For the structuring information, what matters are the predicates which relate two or more NPs.

4.2.4.1 Content

In (27), the content of the subject NP's slot was derived from the extensions of the noun and the VP: it was all individuals that satisfied both. Similarly, the object's slot was all things in the extension of *donkey* which also satisfy the condition of being owned by one of the farmers. In both cases, this can be seen as the intersection of two predicates.

Webber (1979) has proposed that the information contributed by an antecedent may also be all the individuals that the noun describes, independently of the VP, and that properties of the determiner decide which we get. She calls the two classes of determiner intersective and non-intersective. They are illustrated by (28)–(29).

(28) Many farmers own a donkey. They are very cruel.

(29) No farmers own a Rolls Royce. They are too poor.

In (28), *they* refers to the farmers who own a donkey, but in (29), no intersection is taken, and *they* refers to all farmers.

It is not clear that her analysis of the data is correct: Webber uses example (30) to suggest that *few* is a non-intersective determiner, but Evans' (31) seems to show that it can be intersective.

(30) Few linguists smoke, since they (= all linguists) know it causes cancer.

(31) Few MPs came to the party, but they (= MPs who came to the party) had a good time.

I believe a better analysis is to treat at least some determiners as being capable of playing either role, with one role sometimes being preferred over the other. Determiners such as *few* and *no* tend to favour the non-intersective role. (31) is one such example; another is

(32) No men were in the pub. They (= all men) were home watching television.

However, context can force the determiner into the other role, even where the result is unacceptable, as in

(33) No men were in the pub. ?They were having a drink and a sing-song.

A singular anaphor cannot be used in either of these cases. There are cases where the antecedent is non-intersective, but refers to a single thing, perhaps resulting from some contextual factor:

(34) A: Which floor is the bathroom on?
B: There is no bathroom in the house. It (= the contextually relevant bathroom) is in the back yard.

Where the noun phrase is formed from a complex nominal, any other NPs in the nominal also contribute to the anaphoric information. For example in

(35) No farmers who own a donkey beat it. They are kind to them instead.

they refers to all farmers who own a donkey and *them* to their donkeys, with the usual dependence relations.

Determiners behaving non-intersectively seem to have something in common with generics. For example, in

(36) No dodos are alive today. They were hunted to death.

the anaphor seems to refer to dodos in general, but not necessarily to all dodos. Similar remarks can be made for most of the intersective examples above. An analysis in terms of "genera" is put forward by Kamp and Reyle (1990, pp.741-744), although the solution they give is largely a sketch.

Most of the formal development of the following sections will be concerned with determiners purely in their intersective role. However, an adaptation of the theory and a suggestion for choosing which role is to be preferred is presented in section 4.5.1.

There is also an argument for a different division of determiners, into *intersective* and a class I will call *specific*. Specific determiners pass on only part of the intersection, as opposed to all of it. The difference is most apparent with numerals:

(37) Three boys are playing in the park. They have a football.

In (37), *they* could mean either all of the boys who are playing in the park, or some specific three of them even where there are more than three in total. If the numeral is read as *exactly n*, the interpretations coincide, but when it is understood as meaning *at least* both interpretations are available. Another determiner which appears to require the specific interpretation is *a*. If it were taken as intersective, (38) would then be acceptable, which is incorrect.

(38) A man is walking in the park. They (= the men walking in the park) whistle.

Singular *some* is also of this sort. Plural indefinites do generally behave intersectively, as in (39).

(39) Some/several/a few men walk in the park. They (= the men walking in the park) whistle.

The apparently specific behaviour of *a* and singular *some* could be explained by saying that they take the intersection and then, guided by their number properties, take a singular entity from it. In effect, this makes them equivalent to the specific form of *at least one*. What we then need is an explanation of how numerals function on specific/at least readings.

A first possibility is that when the interpretation appears to be specific, some contextual factor limits the domain of quantification, just as *every student passed his exam* does not necessarily make a claim about every student that there is, but about all the ones who are relevant in some circumstances. The specific interpretation can be obtained in a similar way, by saying that the nominal is subject to some contextual limitation, following which the determiner is applied intersectively. In (37), it could be something like *the boys that I can see*. This approach is unsatisfactory in that it leads to a difficult question of its own, namely where the contextual limitation comes from. Another problem created by it is that of early commitment. The specific interpretation of a determiner is indefinite, in that it does not immediately require us to make a commitment to which specific individuals are meant. Subsequent sentences may narrow this down, just as they do with indefinites in general. However, on the contextual limitation account, the decision must be made when the initial sentence is interpreted.

A different solution is to analyse specific determiners as in fact being objects of some other syntactic category. Analyses of this sort have been put forward by Link (1987) and by Verkuyl and van der Does (1991). Without going into details, Link's approach allows numerals to either modify the determiner, for cases such as *at least three children* and *exactly three bananas*, or to function somewhat like adjectives, as in *any two elephants* or *the three men*. Verkuyl and van der Does separate determiners into a specifier, which contains the referential information, and another element containing the quantificational information. Neither Link nor Verkuyl and van der Does make the distinction between intersective and specific determiners, but their general approach of providing more than one place where the numeral and the modifier (*at least*) might appear can be adopted to provide a solution. For the intersective case, we can take *at least three* as being a determiner whose truth conditions are that there are at least three entities satisfying both the nominal and the verb phrase. For the specific case, we analyse the NP as having a null determiner, with the numeral modifying the nominal, restricting it to being made up of the specified number of individuals. The null determiner has truth conditions of existence, and is intersective. Consequently, all determiners may be treated as intersective, with the specific behaviour appearing in the denotation of the modified nominal.

An example may help clarify this. Suppose Joe, Bill, Fred, and Tim are men, and that all but Tim walk in the park. Then

(40) Two^{i} men walk in the park.

on the intersective reading, has truth conditions that at least two of the men walk in the park, and that slot i is set to all the men who do, i.e. Joe, Bill and Fred. On the specific reading, *two men* functions as a nominal, denoting the possible collections of two men such as Joe+Fred, Joe+Bill, Bill+Tim, and so on. For the truth conditions, the null determiner checks that at least one such collection is made up of things that walk in the park, and picks one collection as the content of the slot. One interesting point in this is that the choice is non-deterministic – the two men could be Joe and Bill, or Joe and Fred, or Bill and Fred. As noted above, this leads to alternative contexts.

4.2.4.2 Structure

The evidence of (10)–(11),

(10) Three^i boys bought five^j roses.

(11) They^i took them^j home.

suggested that anaphors are divided into partitions which tend to stay in the same relation as the partitions of the antecedents, and that these partitions are defined by the verb extension. Noun extensions contribute to the structuring only indirectly. In

(41) Two^i couples each own a^j house. They^i live in them^j.

the second sentence requires that the members of each couple have to be considered together in relation to the house they own, in contrast to the less prominent reading where the four people live in the two houses, but apportioned between them in an arbitrary way. This may suggest that the extension of *couples* is contributing to the partitioning. However, it can just as well be explained by saying the first sentence is true only if the pairs of individuals marked as a couple also appear in the extension of *own*. That is, the members of slot i are partitioned in the same way as the extension of *couple* only because *own* must follow the partitioning of *couple* in order for the sentence to be true.

All of the examples of structuring so far have involved transitive verbs, with the structuring information relating one antecedent to another. We might ask whether intransitive verbs also supply structuring information. There are examples to suggest that they do.

(42) Six^i men gathered. They^i shook hands.

The first sentence of (42) can be true if there are two separate occurrences each of three men gathering. In this case, the reading where the three men in each group shook hands with one another but not with members from the other group stands out over the one where the six men, divided up some other way, shook hands. However, it is just as easy to construct a sentence where the structure is not respected:

(42a) They^i arrived in three cars.

I think this suggests that intransitive verbs do not supply structuring information, or perhaps supply it in an indirect way. In (42), the reading of the second sentence may arise because in order to interpret the first one as true, two meeting events are being identified, and the second sentence is interpreted with respect to the events. The connection between the meeting events and the arrival events of (42a) is less close.

It also appears to be the case that the structuring information is only important where there are anaphors to antecedents that have been placed in a dependence relation. Taking (10) again, suppose the discourse continues with

(43) They^i own four^k vases.

Unlike (20)–(21), the reading where the way the three boys are partitioned for rose-buying is the same as the way they are partitioned for vase-owning does not seem more prominent than any other. All that is required is that the content of *they* is the same as the antecedent. Subsequent sentences may force the structuring of *they* in the dependence relation with the vases to be the same as that with the roses. For example,

(44) They^i put the^j roses in the^k vases.

does establish a link between the roses and the vases via their links to slot i. It reads best when each collection of boys that bought some of the three roses put the roses they bought in the vases owned by that collection.

Another source of structuring information is prepositions. An example is:

(45) Some^i people, each with a^j horse, went to the park. They^i rode them^j.

The second sentence is read as saying each person rode the horse they were with. It seems likely (and perhaps not surprising) that structuring information in general is provided by predicates with two or more places.

4.2.4.3 Combined antecedents

Some complications arise with combined antecedents, as in

(46) John^i met Mary^j after work. They went for a walk.

To interpret *they*, we have to take the contents of slot i and combine it with that of slot j This is straightforward enough, but more care must be taken where there is dependence, as in

(47) John^i owns a^j dog and Mary^k a cat^l. They got them from the pet shop.

Now not only do slots i and k have to be merged, but slots j and l have to be also, so that the second sentence is read as each person getting their own animal from the pet shop. Something similar happens with

(48) Two boys buy three roses, and four girls buy five tulips. The children like the flowers.

In merging the slots, the partitions of the original have to be maintained. So if the two boys jointly bought the roses, three girls bought four tulips and one bought one tulip, the merged slots for the children and the flowers would each have three partitions. Merging can become arbitrarily complex: in combining two antecedents, potentially any slots that the antecedents stand in dependence with could be merged.

4.2.5 A formalisation of the model: discourse sets

To formalise the conceptual model, we must first specify a representation of the context. One approach is to derive a number of constraints defining the content of the slots and the dependence relations. For example, from (3),

(3) Every^i farmer owns a^j donkey. They^i beat them^j.

the first sentence yields three constraints, which can be informally expressed as:

1. Slot i contains all farmers who own a donkey.

2. Slot j contains all donkeys owned by a farmer.

3. For each x in slot i, there is a y in slot j such that x owns y; and for each y in slot j there is an x in slot i such that x owns y.

The first two constraints express the content, and the third the structure. To interpret the second sentence, we must make sure that all of the existing constraints are respected, and add an additional one:

4. For each x in slot i, there is a y in slot j such that x beats y; and for each y in slot j there is an x in slot i such that x beats y.

There is some redundancy in the constraints when stated in this way, which can be overcome by combining them into a single, complex constraint.

In a model-theoretic semantics, it is more convenient to work with an "extensional" representation of the context, in which the context is represented as a set of objects (of a sort to be specified) which satisfy the above constraints. Although it is convenient to describe this as if the entire membership of the context were enumerated when the constraints are derived, it is not necessary to approach it in this way. For example, in a computational framework, it may well be more tractable to keep the context in the form of constraints that are evaluated, wholly or partially, only when the slots they define are referred to. Contexts in the representation described here are called *discourse sets*, or DSs.

Each slot of the context is occupied by a number of individuals, divided into partitions. A suitable representation for each slot might therefore be a set of sets of individuals, each "inner" set being one partition. However, I will adopt a representation based on that of Link (1983), so that a slot is a set of sums. The advantage of Link-style sums is that mass nouns can be represented in the same framework as count nouns. There is evidence that just the same mechanism of partitioning is needed for mass nouns; for example

(49) The sugar came in five bags. 10kg of it was in the first bag, and the rest was divided between the remaining four.

Here *the sugar* is a mass noun, and is divided up so that different quantities appear in each partition.[4]

To represent the dependence relations, we can use tuples, each component of the tuple containing one partition in the form of a sum of individuals. To take an example, the context presented as (12),

(12) $$\left\{ \begin{array}{ccc} i & & j \\ b_1 & \leftrightarrow & r_1, r_2 \\ b_2, b_3 & \leftrightarrow & r_3, r_4, r_5 \end{array} \right\}$$

can be represented as the DS

(50) $\{\langle b_1, r_1 \oplus r_2 \rangle, \langle b_2 \oplus b_3, r_3 \oplus r_4 \oplus r_5 \rangle\}$

[4] I will, however, only be covering count nouns in the fragment presented below.

where $\{\ldots\}$ encloses the DS as a whole, $\langle\ldots\rangle$ is a tuple, and $\oplus$ is the Link sum operator.

An "empty" or "null" entity is also needed in some circumstances. Firstly, for sentences such as (20),

(20) Three^i boys who buy five^j roses own four^k vases.

where one slot may be partitioned in more than one way, partitions of one slot which are not linked to anything in one of the other slots are marked by the empty entity in the latter slot. Denoting the empty entity $\perp$, the example context (23), namely

$$(23)\quad \left\{\begin{array}{lclcl} i & & j & & k \\ j & \leftrightarrow & r_1 & & \\ b,f & \leftrightarrow & r_1, r_2, r_3, r_4 & & \\ f & & & \leftrightarrow & v_1 \\ j,b & & & \leftrightarrow & v_2, v_3, v_4 \end{array}\right\}$$

is represented as the following DS.

$$(51)\quad \{\langle j, r_1, \perp\rangle, \langle b \oplus f, r_1 \oplus r_2 \oplus r_3 \oplus r_4, \perp\rangle, \langle f, \perp, v_1\rangle, \langle j \oplus b, \perp, v_2 \oplus v_3 \oplus v_4\rangle\}$$

In extracting information from one or more slots of the DS, we ignore any tuples containing the empty element, since they are relevant only to the dependence relations involving other slots. For example, to interpret (52) following (20)

(52) They^i like the^j roses.

we ignore the latter two tuples of the example DS, since they are relevant only when slots i and k are considered together.

A second reason for needing empty entities is sentences involving *no* or negation, as in (24)–(26). In these cases, there can be slots which are completely empty, which we can represent by having the empty entity appearing in the relevant component of each tuple in the set. For example, in

(26) Some^i men own no^j donkeys.

then a possible DS is

$$(53)\quad \{\langle m_1, \perp\rangle, \langle m_2, \perp\rangle, \ldots\}$$

It was remarked earlier that sentences like this tell us nothing about how slot i is partitioned, since partitioning follows from the dependence relation with the slot j which contains no information in this case. So the following is also a valid DS for (26)

$$(54)\quad \{\langle m_1 \oplus m_2 \oplus \ldots, \perp\rangle\}$$

To interpret

(55) They^i beat them^j.

following (26), the tuples containing the empty entity are dropped, as for (52). As a consequence, it is not possible to assign an interpretation to (55) (without going to a non-intersective or a generic interpretation of *no donkeys*), because there are no members of the DS which contain entities for slot j, corresponding to unacceptability of *them.* It is still perfectly possible to interpret

(56) They^i want to buy some cows instead.

because slot i is non-empty.

There is a third place where empty entities are employed. In

(57) A^i man owns a^j donkey. A^k woman owns a^l horse.

there is no necessary connection between the slots from one sentence and those from the other. Thus, both of the following are possible DSs:

(58) $\{\langle m_1, d_1, \bot, \bot\rangle, \langle \bot, \bot, w_1, h_1\rangle\}$

(59) $\{\langle m_1, d_1, w_1, h_1\rangle\}$

If dependence is subsequently established between, say, slot i and slot k, as in

(60) He^i is married to her^k.

then the first DS will be rejected as a possible context. Unlike the other two uses of empty entities, this one is not essential, in that we could arbitrarily choose to always use the second DS in the above example. If (60) (or any other sentence that establishes a link between slots i and k) never occurs, it does not matter that the man m_1 and the woman w_1 are coincidentally related by appearing in the same tuple: there is nothing that ever uses this information.

There is one potential flaw with this representation of DSs. Taking the example of

(61) Two^i boys buy three^j roses.

on a model where boy 1 buys rose 1, and boy 2 buys roses 2 and 3 together, the DSs could be

(62) $\{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle\}$

But this is not the only possibility. Because we also need to allow the empty entity, any of the following are also allowed:

(63) $\{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle \bot, \bot\rangle\}$
$\{\langle b_1, r_1\rangle, \langle b_1, \bot\rangle, \langle b_2, r_2 \oplus r_3\rangle\}$
$\{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle b_2, \bot\rangle, \langle b_1 \oplus b_2, \bot\rangle, \langle \bot, \bot\rangle\}$
$\{\langle b_1, r_1\rangle, \langle b_1, \bot\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle b_2, \bot\rangle, \langle b_1 \oplus b_2, \bot\rangle\}$
$\{\langle b_1, r_1\rangle, \langle b_1, \bot\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle b_2, \bot\rangle, \langle b_1 \oplus b_2, \bot\rangle, \langle \bot, \bot\rangle\}$

None of these is actually wrong on empirical grounds: the DSs are consistent with the constraints on the information and with the way we have said that information will be extracted from them. They are entirely an artifact of the extensional representation, and in particular the fact that an empty entity must be included. One solution is to nominate one of the equivalent DSs as being canonical and simply not produce the others. The most obvious candidates are the minimal and maximal DSs, i.e. (62) and the last one listed in (63), the former being preferable on grounds of conciseness.

From the point of view of constructing a logic, it does not really matter whether we do this or not, although it implies a need for some care should we ever quantify over DSs. The main concern that having so many equivalent DSs raises is one of computational tractability. The approach I take in the logic of the following section is to allow the full range of equivalent DSs. In section 5.4.2, an algorithm which generates only the minimal DS is sketched.

A final point is that other extensional representations are possible. The DSs presented so far contain tuples with a component for every slot in the context. An alternative representation is to use tuples which have components only for non-empty slots. Each tuple has then to be labelled with the slots it represents. The empty entity can now be dispensed with. For example, the DS shown in (51) can be represented as (64), where subscripts on tuples label the slots which they consist of.

(64) $\{\langle j, r_1\rangle_{i,j}, \langle b \oplus f, r_1 \oplus r_2 \oplus r_3 \oplus r_4\rangle_{i,j}, \langle f, v_1\rangle_{i,k}, \langle j \oplus b, v_2 \oplus v_3 \oplus v_4\rangle_{i,k}\}$

4.2.5.1 Aside: Skolem functions

It has sometimes been suggested that dependence relations could be represented by means of "Skolem" functions, so that in

(65) Every man loves a woman.

the anaphoric contribution of the antecedent *a woman* is a function from men to the women they love. The anaphors in

(66) They prove it to them by giving them flowers.

are interpreted by iterating over the men for *they*, and applying the Skolem function to each such man to interpret *them*. This approach is briefly considered by van den Berg (1991, p.231), who reports that Heim and Scha have both suggested it. He rejects it for formal reasons: Skolem functions do not rest well with the theory of plurals he adopts, and Skolem functions of arbitrary numbers of variables may need to be constructed. There is also an empirical argument, which does not depend in the formal details of the semantic theory. Following (65), we could just as well have

(67) They/the women accept gifts of flowers from them/the men.

To interpret *they*, it would be necessary to iterate over the domain of the Skolem function, and for *them* to work back from the members of the domain to the corresponding member of the range. This suggests that something more like a relation than a function is needed, as DSs provide.

4.2.6 Summary

To recap the discussion of the previous sections: the interpretation of a sentence is carried out with respect to a context, containing one slot for each noun phrase, whether antecedent or anaphor. The result of the interpretation is to place constraints on what constitutes a possible context. More than one context can be consistent with the constraints from a sentence, where there is indeterminacy of reference. For conditional sentences, we construct each possible context for the antecedent sentence, and interpret the consequent sentence with respect to each in turn.

In the formalisation, a context with n slots is represented as a DS, a set of n-tuples. The total content of the slot is equal to the sum of a given component across all the tuples in the set. Each tuple represents partitions that stand in a dependence relation.

The special element $\perp$ may appear in tuples: a tuple such as $\langle f, \perp, \ldots \rangle$ indicates that there is no dependence relation between the f in the first slot and anything in the second slot. The total content of a slot is the intersection of a nominal denotation of a NP and the predicate with which it combines; for example the VP denotation for a subject NP. Dependence relations are defined by the predicate relating two or more NPs.

Although this description has a very "extensional" flavour, DSs can equally be thought of in terms of collection of constraints on the content of slots and the relations between them. The constraints must be respected whenever reference is made to the slots. For example, in

(68) Every^i farmer owns a^j donkey. He^i beats it^j. He^i is cruel to it^j.

the interpretation of the third sentence must respect four constraints: that the content of slot i consists of all farmers who own a donkey, and of slot j of all donkeys which are owned by a farmer, and that the partitions of slots i and j must stand in both the *own* and *beat* relations.

In the next section, I present a logic which interprets formulae derived compositionally from English sentences to yield their truth conditions and the constraints they place on possible contexts. The evaluation of the TAI and an exploration of how the conceptual model relates to DRT and dynamic logics appears in chapter 5.

4.3 L(GQA): a logic for the DS model

L(GQA) – logic of generalized quantifiers with anaphora – is a formalisation of the DS model. L(GQA) formulae are derived from natural language sentences, and interpreted through two interpretation functions. One supplies the truth conditions, and the other, called the anaphoric interpretation, gives the conditions for a DS to be correct. The development of the logic in this section proceeds by defining the functions on L(GQA) expressions derived from sentences not containing anaphors, and then adds the terms needed for anaphors. The discourse level part of the logic is added on top of L(GQA) in section 4.3.5. Conditionals and a sequencing operator are included at this stage, the resulting logic being called L(GQAD).

The model of acceptability is the one described in section 2.4.2, in which truth conditions and acceptability conditions are not separated. In L(GQA) this takes the form of making the overall interpretation of a closed formula, derived from a sentence, be the conjunction of the truth conditional and anaphoric parts of the interpretation. The merits of this approach will be considered in chapter 5; one of the roles of L(GQA) is to test out the idea.

The procedure for translating English sentences into formulae of L(GQA) forms a significant part of the empirical content of the theory. It is described in section 4.4. I will use the term translation for the L(GQA) expression derived from an English one in preference to representation.

The semantics as presented here treats all determiners as intersective. As discussed in section 4.2.4.1, there are also non-intersective uses of determiners. A proposed treatment of such determiners appears in section 4.5.1.

4.3.1 Preliminaries

L(GQA) draws on the language of generalized quantifiers L(GQ) described in Barwise and Cooper (1981) for some parts of its syntax and interpretation, and on the logic of plurals and generalized quantifiers LP+GQ of Link (1987). Here I give a brief introduction to the theories.

4.3.1.1 Generalized quantifiers

The representation of noun phrases as generalized quantifiers (GQs) originates in the work of Montague. Barwise and Cooper (1981) study the formal properties of GQs and their linguistic significance; also see van Eijck (1988), van Benthem (1989) and Westerståhl (1989) for further investigations. A generalized quantifier is a set of sets of entities. Considered as the denotation of a noun phrase, this can be thought of as the set of properties which the noun phrase has, i.e. the set of all 1-place predicates which it satisfies. Determiner denotations are functions from the nominal denotation, a set, to a GQ.

In a simple sentence of the form *NP VP*, Barwise and Cooper take the VP denotation as being a set of individuals, and the sentence as being true if this set is a member of the generalized quantifier for the NP, $\|VP\| \in \|NP\|$. Some sample GQs are:

every man: $\{X \subseteq E \mid man' \subseteq X\}$, i.e. the set of sets containing all men. E is the domain of individuals.

some man: $\{X \subseteq E \mid man' \cap X \neq \emptyset\}$, i.e. the set of sets containing at least one man.

most men: $\{X \subseteq E \mid card(X \cap man') \geq card(X - man')\}$, i.e. the set of sets containing at least half the members of the set of men ($card(S)$ is the cardinality of set S).

John: $\{X \subseteq E \mid john \in X\}$, i.e. the set of sets containing the individual *john*.

Barwise and Cooper propose a number of universals, properties which are shared by all GQs occurring in natural language. The three most important are what van Eijck (1988) calls *extension*, *conservativity*, and *isomorphy*. Using the notation $D_E AB$ to mean a determiner D over domain E applied to the sets A and B, equivalent to $B \in D_E A$, the properties are defined by van Eijck as follows:

Extension. A determiner D satisfies the condition of extension if for all $A, B \subseteq E \subseteq E'$: $D_E AB \Leftrightarrow D_{E'} AB$. That is, determiners are stable under growth and contraction of the universe, provided the universe contains all of A and B.

Conservativity (called the *live-on* property by Barwise and Cooper). A determiner D satisfies the condition of conservativity if for all $A, B \subseteq E$: $D_E AB \Leftrightarrow D_E A(A \cap B)$. So the part of B which is not within A is irrelevant for evaluating the truth value of the quantification.

Isomorphy. A determiner D satisfies the condition of isomorphy if, given a bijection f from E to E', then $D_E AB \Rightarrow D_{E'} f[A] f[B]$. Hence the membership of A and B does not matter, only their cardinalities.

It follows that a determiner over the sets A and B can generally be expressed as a relation between the cardinalities of A and $A \cap B$, provided the domains are finite. Some examples, expressed in this form, are:

every: $\lambda mn.m = n$

most: $\lambda mn.n \geq m/2$

few: $\lambda mn.n \leq f(m)$, for some measure of "fewness" f.

some: $\lambda mn.n \geq 1$

no: $\lambda mn.n = 0$

For example, the truth conditions of

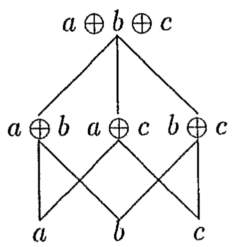
(69) Most men whistle.

are $card(man' \cap whistle') \geq card(man')/2$. That is, the number of men who whistle is at least half of the number of men in total. Some possible exceptions to these universals have been suggested. In general they can be explained either as being a different syntactic construction, or as requiring a quantifier which is broadly similar in principle, but of a higher type; van Eijck (1988) is again a good reference.

In L(GQA), I adopt a generalized quantifier approach, but with one major modification. Most GQ theory has been concerned only with distributive readings of sentences, for which a domain E of individuals is sufficient. To cover collective readings, we will need to be able to quantify over plural entities, such as collections of individuals. The details appear below.

4.3.1.2 Link's theory of plurals

The theory of plurals developed by Godehard Link provides the other ingredient for obtaining collective readings. For the first exposition of the theory, see Link (1983). A generalized quantifier version appears in Link (1987), and some other refinements can be found in Link (1984a), Link (1984b) and Landman (1987). Rather than represent plural entities as the set of individuals which make up the entity, Link uses a sum operator $\oplus$, which forms plural entities out of smaller ones. As mentioned in section 4.2.5, the two representations are equivalent for count nouns. An inclusion relation, analogous to $\subseteq$, can be defined, and the resulting partial order forms a complete join-semilattice. As an example, on a domain containing three individuals a, b and c, the lattice is (with elements including the ones below them):



Link's approach is preferable for mass nouns, as in the example quoted earlier:

(49) The sugar came in five bags. 10kg of it was in the first bag, and the rest was divided between the remaining four.

The same lattice structure can be used, but with the difference that the entities in it cannot be reduced to individuals. In theory, at least, we can keep on dividing up each mass quantity into smaller and smaller portions.

Both Link and Landman (1987) show that a notion of *groups* is also needed for a fully adequate treatment of plurals. Roughly, groups are entities which are normally indivisible, but which can be broken down further if an additional operator is applied. An example (from Landman) which shows the need for them is

(70a) Committee A paid an official visit to South Africa.

(70b) Committee B paid an official visit to South Africa.

Sentence (70a) can be true and (70b) false even if the committees have the same members. Each committee is represented as a group, and although the individuals comprising the groups are the same, they are distinct entities. The groups can be broken down when the need arises:

(71) They/The members of the committee received death threats.

The extra complications involved in groups will be omitted, at the cost of some limitation to the coverage. The ideas behind the logic should extend to groups.

4.3.2 L(GQA): first version

The syntax of L(GQA), which has some intentional similarities with Barwise and Cooper's L(GQ), is as follows.

Basic expressions:

1. Determiners, symbolised as D below. Examples: *most*, *few*, *every*, *no*, *the*, Φ and for each integer $n \geq 1$, the symbols $n_=$, $n_\geq$ and $n_\leq$.
2. Variables: $x, y, \ldots$.
3. Predicate symbols: R. Examples: *buy*, *walk*, *farmer*, *own*, *beat*, *john*, *mary*. Each predicate has a specified number of argument places.
4. Indices: i, j.
5. Number terms: n, where n is a natural number.

Derived expressions:

1. Quantifiers: Q.
2. Set terms: S.
3. Formulae: F.

Formation rules:

F1. If Q is a quantifier, i is an index and S is a set term, $Q(i)S$ is a formula.

F2. If R is an n-place predicate symbol, and $x_1, \ldots, x_n$ are variables, $R(x_1, \ldots, x_n)$ is a formula ($n \geq 2$).

F3. If D is a determiner and S a set term, DS is a quantifier.

F4. If x is a variable and F is a formula, $\hat{x}[F]$ is a set term.

F5. If R is a predicate symbol, R_s is a set term.

F6. If S_1 and S_2 are set terms, $S_1 \bigwedge S_2$ is a set term.

F7. If n in a natural number and S is a set term, $n(S)$ is a set term.

The subscript on R_s is used to make clear the type of the expression. Φ is the null determiner. The symbol is used in preference to Link's $\emptyset_E$, to avoid confusion with the meta-language empty set symbol. Set terms of the form $\hat{x}[F]$ stand for the set of all x for which F holds. Rule F6 forms the intersection of two set terms, and F7 finds the members of the S of size n.

A closed formula is one in which each variable x appears only within the body of a complex set term $\hat{x}[F]$. The translation of a natural language sentence is a closed formula. There is no implication operator on L(GQA): using generalized quantifiers takes away the need for it in universal quantification, and conditionals are handled by the discourse interpretation, described in section 4.3.5. Indices will play no role until the anaphoric interpretation is introduced. The initial version of L(GQA) does not contain the logical connectives of negation, conjunction and disjunction. They are added in a later section.

The conceptual model of the previous section gave every NP its own index, with the contents of slots being equated when an anaphor was paired with an antecedent. I will make the minor simplification of using the same index, as in the examples of section 4.2.

4.3.2.1 Top-level interpretation

L(GQA) has two interpretation functions, written $[\![.]\!]_t{}^{W,m}$ and $[\![.]\!]_a{}^{W,m}$. The truth conditional part, $[\![.]\!]_t{}^{W,m}$, interprets expressions with respect to a model (suppressed), a DS W and a map m, which keeps track of bindings to variables, as described in subsequent sections. For a formula F, the truth conditional part of the interpretation specifies the conditions on a model for F to be true, as in many semantic theories. The DS parameter is used only for anaphoric expressions, where it makes the anaphoric information available.

The anaphoric part, $[\![.]\!]_a{}^{W,m}$, specifies the conditions which the DS must meet in order to accurately reflect the anaphoric properties of antecedents and anaphors represented in F. In effect, it expresses the constraints on DSs described in section 4.2.5. The interpretation process can be understood by treating the model as fixed, and the anaphoric interpretation as an instruction to set up the correct DS given the predicate extensions in the model.

Given a closed formula F of L(GQA), a top-level interpretation which combines the two parts is also defined:

$$[\![F]\!]^W \;=\; [\![F]\!]_t{}^{W,[\,]} \wedge [\![F]\!]_a{}^{W,[\,]}$$

$[\,]$ is the empty map, and $\wedge$ is Boolean conjunction. This means that the interpretation of F is true with respect to a model and a DS, if the model meets the conditions placed on predicates appearing in F, and if the DS fits the anaphoric properties of F. As already noted, this follows the model of acceptability and truth presented in section 2.4.2.

4.3.2.2 Model and meta-language

The model for L(GQA) is a triple $\langle \mathcal{E}, F, I \rangle$, where $\mathcal{E}$ is a domain of atoms, F is a function mapping each predicate symbol to its extension, and I a domain of indices. Atoms are model-theoretic entities which cannot be expressed as a sum of other entities, and represent individuals. Model-theoretic conditions are expressed in a meta-language containing standard set and truth operators, together with $\oplus$, $\downarrow$ and $\langle . \rangle$, which are defined below. *card*(.) means set cardinality. A number of additional operators are derived from the basic ones. The definitions of the meta-language operators are also listed in appendix B.

Sums and domains

The sum operators are based on those of Link (1983), as described above. The structure differs from Link's in adding a bottom element $\perp$ which is needed to represent the empty entity in DSs. The resulting structure is a complete lattice.

The domain of atoms including $\perp$ is denoted $\mathcal{E}_\perp = \mathcal{E} \cup \{\perp\}$. The join operator is Link's $\oplus$. $\oplus$ meets the conditions of commutativity ($a \oplus b = b \oplus a$), associativity ($(a \oplus b) \oplus c = a \oplus (b \oplus c)$) and idempotence ($a \oplus a = a$). $\perp$ is defined by having the property $\forall x \in \mathcal{E}_\perp[x \oplus \perp = x]$. $\perp$ acts like the empty set in a powerset model of plurals. $\oplus$ may be applied to sums, written $a \oplus b$, and to sets of sums, written $\oplus A$, the latter meaning the sum of all the members of A. $\oplus(\emptyset) = \perp$. The closure of $\mathcal{E}$ under $\oplus$ is denoted $\mathcal{D}$, with $\mathcal{D}_\perp = \mathcal{D} \cup \{\perp\}$. The product domains $\mathcal{E} \times \mathcal{E} \ldots \times \mathcal{E}$ (n times) is written as $\mathcal{E}^n$, and similarly for $\mathcal{D}^n$.

The operators on sums are, where A is a subset of $\mathcal{D}_\perp$:

$Ats(a)$ is the set of atoms A such that $\oplus A = a$. $Ats(\perp) = \emptyset$.

$Crd(a)$ is the sum-cardinality, i.e. number of atoms making up a sum. $Crd(a) = card(Ats(a))$.

$Sup(A)$ is the supremum of A, i.e. the unique $a \in A$ such that

$$\forall a' \in A[(a \neq a') \rightarrow Crd(a') < Crd(a)]$$

It is not defined if there is more than one $a \in A$ meeting the condition. $Sup(A)$ is always defined if A is closed under summation.

Tuples

The n-tuple of $a_1, \ldots, a_n$ is written $\langle a_1, \ldots, a_n \rangle$, with 1-tuples $\langle a \rangle$ abbreviated to a. To avoid confusion with members of sets, the term *component* is used to refer to a part of a tuple. The basic operator on tuples is extraction $\downarrow$ which forms a new tuple of the specified components:

$\langle a_1, \ldots, a_n \rangle \downarrow \langle i_1, \ldots, i_k \rangle = \langle a_{i_1}, \ldots, a_{i_k} \rangle$. Defined only if, for each i_j, $1 \leq i_j \leq n$.

Extraction can also be applied to sets of tuples.

$$W \downarrow \langle i_1, \ldots, i_n \rangle = \{w \downarrow \langle i_1, \ldots, i_n \rangle \mid w \in W\}$$

We will often want to extract some components from a set and sum the result, for which $\Downarrow$ is used:

$$W \Downarrow \langle i_1, \ldots, i_n \rangle = \oplus(W \downarrow \langle i_1, \ldots, i_n \rangle)$$

The sum operator on tuples acts component by component. It is defined only on sets where all the tuples have the same number of components:

$$\oplus\{\langle a_1, \ldots, a_n \rangle, \ldots, \langle z_1, \ldots, z_n \rangle\} = \langle \oplus\{a_1, \ldots, z_1\}, \ldots, \oplus\{a_n, \ldots, z_n\} \rangle$$

The restrictor operator $W \backslash (X, i)$ finds all members of the set W having X in component i.

$$W \backslash (X, i) = \{w \in W \mid w \downarrow i = X\}$$

Note that $W\backslash(X,i)\Downarrow i$ is equal to X, since $W\backslash(X,i)$ is the subset of W containing X in the i component. This is true even for $X = \bot$. The identity is occasionally useful in understanding and simplifying formulae. To take an example, consider the set

$$W = \{\langle a_1, b_1, c_1\rangle, \langle a_1, b_2, c_2\rangle, \langle a_3, b_3, c_3\rangle\}$$

where the components of the tuples represent slots i, j and k. The extraction of component j is:

$$W\downarrow j = \{b_1, b_2, b_3\}$$

and hence $W\Downarrow j = b_1 \oplus b_2 \oplus b_3$. An example of restriction is $W\backslash(a_1, i)$, the part of W having a_1 in slot i:

$$W\backslash(a_1, i) = \{\langle a_1, b_1, c_1\rangle, \langle a_1, b_2, c_2\rangle\}$$

and hence $W\backslash(a_1, i)\Downarrow i = a_1 \oplus a_1 = a_1$.

Sets and predicates

For an n-ary predicate symbol R, $F(R)$ is a set of n-tuples of sums, called the *essential extension* of R, and containing all tuples that satisfy the predicate in a genuine rather than a contingent way. In the case of nouns, this means reducing the extension of the noun to the smallest elements which can still be said to be described by the noun, and for verbs similarly, to the minimal events, actions, states and so on described by the verb. Taking the predicate *lift* as an example, suppose that Joe and Bill lift table t_1 together, but do not lift it individually, and Fred lifts the same table alone. The essential extension $F(\mathit{lift})$ contains the tuples $\langle j \oplus b, t_1\rangle$ and $\langle f, t_1\rangle$, but it does not contain $\langle j \oplus b \oplus f, t_1\rangle$, even though it is true that all three men do lift the table. If Fred had lifted t_2, the essential extension would contain $\langle f, t_2\rangle$ but not $\langle j \oplus b \oplus f, t_1 \oplus t_2\rangle$. For predicate symbols which correspond to proper nouns, such as *john*, $F(P)$ will typically have only one member, namely the atom corresponding to John. For natural numbers n, $F(n)$ is the set of sums of the given size, i.e. $\{x \in D \mid \mathit{Crd}(x) = n\}$. In all three cases, $F(P)$ is written as P' for convenience. It is stipulated that for no P, $\bot \in F(P)$.

There are three special predicates in the model, which are used for semantic number agreement on anaphors: *sg* for singular objects, *pl* for plural objects, and *un* for objects of unspecified number. The extensions are equal to $\mathcal{E}$, $\mathcal{D} - \mathcal{E}$ and $\mathcal{D}$, respectively.

For sets of n-tuples S, the closure under summation, $^*S$, is defined by:

$$^*S = \{Y \in \mathcal{D}^n \mid \exists X \subseteq S[Y = \oplus X]\}$$

$^*S$ contains all sums that can be formed from one or more members of S. In effect, this adds all contingent members to the set. From the definition, it follows that the intersection of sum closed sets is also sum closed. For basic predicates R, $^*R'$ is usually written in the slightly neater form R^*. For a set of 1-tuples, the set of all sums which can be formed by breaking down any sums and summing the resulting atoms is written $\sharp(S)$, defined by

$$\sharp(S) = {}^*(\mathrm{Ats}(\oplus S))$$

A set may be applied to one or more sums. If S is a set of n-tuples, and $x_1, \ldots, x_n$ are sums, then

$$S(x_1, \ldots, x_n) = \langle x_1, \ldots, x_n\rangle \in S$$

In the anaphoric interpretation, we will also need a special case which allows arguments to be $\perp$, for which the notation $S_\perp$ is employed.

$$S_\perp(x_1,\ldots,x_n) \;=\; (\langle x_1,\ldots,x_n\rangle \in S) \vee (x_1\!=\!\perp) \vee \ldots \vee (x_n\!=\!\perp)$$

Notational simplifications of $\{X \in \mathcal{D} \mid S(\ldots,X)\}$ to $S(\ldots)$, and of $\{X \in \mathcal{D}_\perp \mid S_\perp(\ldots,X)\}$ to $S_\perp(\ldots)$ will be used.

A set S is distributive if the atoms of every sum are present:

$$Distr(S) = \forall s \in S[Ats(s) \subseteq S]$$

On this definition, the empty set is distributive.

The operator *Max* finds the largest members of a set S, i.e. those sums in S such that there are no members of S made up of more atoms, or $\{\perp\}$ when S is empty.

$$Max(S) = \begin{cases} \{s \in S \mid \forall t \in S[Crd(t) \leq Crd(s)]\} & \text{if } S \neq \emptyset \\ \{\perp\} & \text{otherwise} \end{cases}$$

Max_E is the same as *Max*, provided its argument is distributive, in the sense that the atoms of every sum are present in the set. Otherwise, the result is the empty set.

$$Max_E(S) = \begin{cases} Max(S) & \text{if } Distr(S) \\ \emptyset & \text{otherwise} \end{cases}$$

Note that if S is closed under summation, as will often be the case, then $Max(S)$ contains only a single member, consisting of a sum made up of all atoms that appear somewhere in S, whether actually as atoms, or as a part of a sum. This is equal to $Sup(S)$.

Extensions of n-place predicates are subsets of $\mathcal{D}^n$, and DSs subsets of the product domain $\mathcal{D}_\perp \times \mathcal{D}_\perp \times \ldots$ (to an arbitrary number of terms).

4.3.2.3 Interpretation

The truth conditional interpretation of L(GQA) is defined recursively by the following functions. W is a DS parameter, which will be needed when anaphoric expressions are added. The parameter m is a map which keeps track of bindings of L(GQA) variables to meta-language variables and indices. It has no significance other than as a technical device, i.e. it does not represent persistent information. $m[x/\langle X, i\rangle]$ means the map which is identical to m except that it binds x to X and i. $m_1(x)$ yields X; $m_2(x)$ yields i. Interpretation of a closed formula starts with the empty map [ ]. The model is suppressed.

Lambdas are used to pass meta-language arguments to the interpretation functions.

1. $[\![Q(i)S]\!]_t^{W,m} \ = \ [\![S]\!]_t^{W,m} i \in [\![Q]\!]_t^{W,m} i$
2. $[\![R(x_1, \ldots, x_n)]\!]_t^{W,m} \ = \ [\![R]\!]_t^{W,m}(m_1(x_1), \ldots, m_1(x_n))$
3. $[\![DS]\!]_t^{W,m} \ = \ \lambda i.([\![D]\!]_t^{W,m} i)([\![S]\!]_t^{W,m} i)$
4. $[\![\hat{x}[F]]\!]_t^{W,m} \ = \ \lambda i.{}^*\{X \in \mathcal{D} \mid [\![F]\!]_t^{W,m[x/\langle X,i\rangle]}\}$
5. $[\![R_s]\!]_t^{W,m} \ = \ \lambda i.[\![R]\!]_t^{W,m}$
6. $[\![S_1 \bigwedge S_2]\!]_t^{W,m} \ = \ \lambda i.([\![S_1]\!]_t^{W,m} i) \cap ([\![S_2]\!]_t^{W,m} i)$
7. $[\![n(S)]\!]_t^{W,m} \ = \ \lambda i.n' \cap ([\![S]\!]_t^{W,m} i)$
8. $[\![R]\!]_t^{W,m} \ = \ R^*$
9. $[\![D]\!]_t^{W,m} \ = \ \lambda i \lambda p.\{X \subseteq \mathcal{D} \mid D'(p, X)\}$, where D' is defined below.

From the above definitions, all set extensions are closed under summation, except for those formed with numerals. As a consequence, the supremum can be guaranteed to exist in such cases. When a numeral is involved, in expressions such as $n(S)$, the resulting set has members which are all of the same size.

4.3.2.4 Determiners and quantifiers

As in generalized quantifier theory, determiners are expressed as a relation between two sets p and q. Some determiners are classed as distributive, meaning that they can only apply to distributive set terms; the remaining determiners can apply to any sort of set term. For distributive determiners, the definitions can be formulated using exactly the same relations as the standard definitions, but first reducing the set terms to their atoms. The sum cardinality of the supremums could be used just as well, i.e. $Crd(Sup(p))$ and $Crd(Sup(p \cap q))$. Link (1987) uses the supremum-cardinality definition in the generalized quantifier version of his theory of plurals, and comments that the same definitions can be used with mass terms, if a measure function is used in place of $Crd(\ldots)$.

The definitions given in section 4.3.1 may then be used. They are repeated here, expressed as functions which are applied to $card(p \cap \mathcal{E})$ and $card(p \cap q \cap \mathcal{E})$. Intersection with $\mathcal{E}$ reduces the set to its atomic part. For distributive sets, no information is lost in doing so.

every: $\lambda mn.m = n$

most: $\lambda mn.n \geq m/2$

few: $\lambda mn.n \leq f(m)$, for some measure of "fewness" f.

some: $\lambda mn.n \geq 1$

no: $\lambda mn.n = 0$

The definitions of *some'* and *no'* are interesting, in that they can be restated in terms which do not rely on the distributivity, consistent with them having collective readings.

$$some'(p, q) = p \cap q \neq \emptyset$$

$$no'(p, q) = p \cap q = \emptyset$$

some' as it stands does not distinguish between the singular and plural uses of the determiner. For a strictly plural use, the (cardinality) definition is $\lambda mn.n > 1$. *All* may also be used collectively, with the definition being

$$all'(p, q) = Sup(p) \in q$$

An alternative definition is $Max(p) \subseteq q$, which does not rely on the existence of the supremum. This may be needed for NPs such as *all three men*; see Link (1987) for a discussion of NPs of this sort.

For each n, there are determiners representing *at least* and *at most*:

$$n'_{\geq}(p, q) = \exists x \in p \cap q[Crd(x) \geq n]$$

$$n'_{\leq}(p, q) = \exists x \in p \cap q[Crd(x) \leq n]$$

The *exactly* determiner corresponds to the use of the word to mean that the specified number of things satisfy the condition and that there is no larger collection that satisfies the condition.

$$n'_{=}(p, q) = \exists x \in Max(p \cap q)[Crd(x) = n]$$

There is also an English use of the word to mean, roughly, a specific collection with exactly the given number of members, for which the null determiner Φ with a set term such as $3(man_s)$ is used. The definition could also have been stated as $\forall x \in Max(\ldots)$, since all members of the Max of a set have the same sum-cardinality. As already noted, if $p \cap q$ is closed under summation (which is true if p and q are), $Max(p \cap q)$ contains a single member. In this case, the definition $Crd(Sup(p \cap q)) = n$ is equivalent.

The null determiner has truth conditions of existence, and is identical to collective *some*:

$$\Phi'(p, q) = p \cap q \neq \emptyset$$

A difference between $(\Phi\ 3(man_s))$ and $(3_{\geq}\ man_s)$ is that the former guarantees there is a collection of three men, and the latter asserts the existence of a collection containing three or more men, but does not require that there is a collection of exactly three.

The definite determiner ensures that the whole of p satisfies q:

$$the'(p, q) = Sup(p) \in q$$

Singular and plural versions could easily be formulated.

In the interests of conciseness, I will generally replace $q \in \{X \subseteq \mathcal{D} \mid D'(p, X)\}$ by the equivalent $D'(p, q)$. All of the determiners satisfy the conservativity property. This is clearly so for all definitions other than *all* and *the*, since the definitions are expressed in terms of p and $p \cap q$, but not q alone. The remaining definitions all contain q only in the expression $Sup(p) \in q$. Since $Sup(p) \in p$, provided $Sup(p)$ is defined,

$$Sup(p) \in q = Sup(p) \in q \wedge Sup(p) \in p = Sup(p) \in p \cap q$$

from which conservativity follows.

4.3.2.5 Examples

(72) receives the translation (72a) with interpretation (72.T). Indices appear as superscripts on determiners in the sentences; since they are associated with quantifiers, they should be thought of as being annotations on NP nodes.

(72) Most^i donkeys bray.

(72a) $(most\ donkey_s)(i)\ bray_s$

(72.T) $most'(donkey^*, bray^*)$

The result of expanding out the determiner definition is

$$card(donkey' \cap bray') \geq card(donkey')/2$$

That is, the sum total of donkeys that bray is at least half the size of the sum total of donkeys.

(73) receives the translation (73a) with interpretation (73.T), both quantifiers being taken as "exactly", $n_=$.

(73) Two^i boys buy three^j roses.

(73a) $(2_=\ boy_s)(i)\ \hat{x}[(3_=\ rose_s)(j)\ \hat{y}[buy(x,y)]]$

(73.T) $2'_=(boy^*, {}^*\{X \in \mathcal{D} \mid 3'_=(rose^*, buy^*(X))\})$

The set $\{X \in \mathcal{D} \ldots\}$ consists of all those sums X for which there are three roses collectively bought by X. To satisfy (73.T), there must be a sum of boys of size 2 equal to a sum of such X's, i.e. which can be subdivided into sums each buying three roses.

To see how (73) works in detail, suppose we have the following model. b_1, b_2 and b_3 are boys, c is some individual which is not a boy, and r_1, r_2, r_3 and r_4 are roses. b_1 buys r_1; b_2 buys both r_2 and r_3 together; c buys r_4. The basic extensions are:

$boy' = \{b_1, b_2, b_3\}$

$rose' = \{r_1, r_2, r_3, r_4\}$

$buy' = \{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle c, r_4\rangle\}$

The extensions closed under summation are:

$boy^* = \{b_1, b_2, b_3, b_1 \oplus b_2, b_1 \oplus b_3, b_2 \oplus b_3, b_1 \oplus b_2 \oplus b_3\}$

$rose^* = \{r_1, r_2, r_3, r_4, r_1 \oplus r_2, r_1 \oplus r_3, r_1 \oplus r_4, r_2 \oplus r_3, r_2 \oplus r_4, r_3 \oplus r_4, r_1 \oplus r_2 \oplus r_3,$
$r_1 \oplus r_2 \oplus r_4, r_2 \oplus r_3 \oplus r_4, r_1 \oplus r_2 \oplus r_3 \oplus r_4\}$

$buy^* = \{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle c, r_4\rangle, \langle b_1 \oplus b_2, r_1 \oplus r_2 \oplus r_3\rangle, \langle b_1 \oplus c, r_1 \oplus r_4\rangle, \langle b_2 \oplus c, r_2 \oplus r_3 \oplus r_4\rangle,$
$\langle b_1 \oplus b_2 \oplus c, r_1 \oplus r_2 \oplus r_3 \oplus r_4\rangle\}$

Consequently, there are two values of X for which $3'_=(rose^*, buy^*(X))$ holds, namely $b_1 \oplus b_2$ and $b_2 \oplus c$. Hence we have

$${}^*\{X \in \mathcal{D} \mid 3'_=(rose^*, buy^*(X))\} \;=\; {}^*\{b_1 \oplus b_2, b_2 \oplus c\} \;=\; \{b_1 \oplus b_2, b_2 \oplus c, b_1 \oplus b_2 \oplus c\}$$

The truth condition (73.T) holds, since the intersection of this set with boy^* is $\{b_1 \oplus b_2\}$, the sole member of which, $b_1 \oplus b_2$, has sum-cardinality of two, in keeping with the determiner definition. Note that this is the cumulative reading.

4.3.2.6 Sentence readings

L(GQA) follows and extends the work of Verkuyl and van der Does (1991) described in section 2.6, in being vague between the D, C_1 and C_2 (distributive and collective) readings, and also in obtaining the cumulative reading using the same semantics. The same set terms are used for all three cases. Basic set terms are closed under summation so that, for example, the sentence

(74) Six men met.

will be true if there is a sum of four men who met and another sum of two men who did so as well. For similar reasons, complex set terms $\hat{x}[F]$ are sum closed. In the C_2 example given for

(75) Four men lifted three tables

in section 2.6, the extension of $\hat{x}[(3_= \; table)(j) \; \hat{y}[lift(x,y)]]$ includes $m_1 \oplus m_2, m_3$ and m_4, and all sums formed from them, so there is a sum which consists of four men. Numeral set terms, n in $n(S)$, are not sum closed, to allow them to represent the set of objects of exactly the specified size.

To adopt an ambiguity account rather than a vagueness one, the logic would need some means of indicating which reading of a set term was intended. The interpretation is changed so that D (distributive) readings restrict set term extensions to be the sum closure of subsets of $\mathcal{E}$, meaning every atom does what the set term specifies. Example: distributive *three men each walk* would be true if there was a sum of three men, each atom of which walked. C_1 readings still draw on domain $\mathcal{D}$, but VP extensions are not sum closed, so their extension consists of all things that do the specified action together. Example: *three boys gather* would be true if there was a sum of three boys, which was in the essential extension of *gather*. C_2 readings are unchanged.

Closing set terms under summation also means that the interpretation of (75) yields Scha's cumulative reading when the quantifiers are taken as "exactly". To illustrate, take the model corresponding to the diagram (152a) in section 2.6. The interpretation of (75) is:

(75.T) $4'_=(man^*, {}^*\{X \in \mathcal{D} \mid 3'_=(table^*, lift^*(X))\})$

equivalent to

$$\exists u \in Max(man^* \cap {}^*\{X \in \mathcal{D} \mid \exists v \in Max(table^* \cap lift^*(X))[Crd(v)=3]\})[Crd(u)=4]$$

The essential extension of *lift* is

$$\{\langle m_1, t_1\rangle, \langle m_2, t_2\rangle, \langle m_3 \oplus m_4, t_3\rangle, \ldots\}$$

giving a sum closed extension of which $\langle m_1 \oplus m_2 \oplus m_3 \oplus m_4, t_1 \oplus t_2 \oplus t_3\rangle$ is a member. It follows that the sum $m_1 \oplus m_2 \oplus m_3 \oplus m_4$ is in the set ${}^*\{X \in \mathcal{D} \mid \ldots\}$, and (75.T) holds.

On the basis of similar reasoning, the sentence

(76) Four boys buy three roses.

will be interpreted as true in a model such as:

b_1 buys r_1

b_2 buys $r_2 \oplus r_3$

b_3 buys r_4

b_4 buys $r_5 \oplus r_6$

Although this is not an obvious reading of the sentence, I think it is a possible one; consider a situation where the first two boys went into a shop and came out with three roses, and the second two boys did the same.

4.3.2.7 Proper nouns

Proper nouns are translated as the definite determiner *the* applied to a predicate symbol for the proper noun. Thus, *Fred* translates into the L(GQA) expression (*the fred*$_s$), "the thing which is Fred". Taking the extension *fred'* as the set containing the atom f, the interpretation is

$$\{X \subseteq \mathcal{D} \mid the'(fred^*, X)\} = \{X \subseteq \mathcal{D} \mid f \in X\}$$

For proper nouns that denote sets, the usual range of distributive and collective readings is possible. (77) can mean either that all of the Tebbits ate a cake together, or that they each ate a different one, and so on.

(77) The Tebbits ate a cake.

In this example, *The Tebbits* is treated as a single lexical item, rather than as the determiner applied to a proper noun *Tebbits*.

One minor problem with this analysis is that there can be more than one bearer of a proper noun. *Fred* can mean various different people answering to that name, and there is more than one group of people called *The Tebbits*. I will make the simplifying assumption that there is a different predicate symbol for each bearer of the name. An alternative would be to make the translation of a proper noun P be (*some* P) (or, equivalently, (Φ P)), that is, some bearer of the proper name.

4.3.2.8 Partitions and covers

Four different notions of collectivity were explored in section 2.6.1. In the truth conditional part of L(GQA), set term extensions are formed by sum closing sets, which corresponds to finding a covering. To incorporate a more restricted notion into L(GQA), the definition of sum closure for predicates and complex set terms would be

$$^*S = \{Y \in \mathcal{D} \mid \exists X \subseteq S[Y = \oplus X \wedge P(X,Y)]\}$$

with $P(X,Y)$ as follows:

1. Pseudo-partition: $card(X) \leq Crd(Y)$.
2. Minimal cover: $\neg\exists X' \subset X[\oplus X' = Y]$.
3. Partition: $\forall X_i, X_j \in X[Ats(X_i) \cap Ats(X_j) \neq \emptyset]$

As van der Does notes, it is not completely clear which notion is the correct one. The best candidate appears to be pseudo-partitioning, but only for transitive verbs. To carry this out in L(GQA), we would define sum closure using P for pseudo-partitioning for complex set terms, and use the standard definition of sum closure in section 4.3.2.2 otherwise.

4.3.3 L(GQA) with DSs

There is a second interpretation function for L(GQA), which produces the conditions for a DS to contain the correct information, called the anaphoric conditions. The definition of *res* follows.

1. $[\![Q(i)S]\!]_a{}^{W,m} \;=\; [\![Q]\!]_a{}^{W,m} i([\![S]\!]_t{}^{W,m} i)([\![S]\!]_a{}^{W,m} i)$
2. $[\![R(x_1,\ldots,x_n)]\!]_a{}^{W,m} \;=\; [\![R]\!]_a{}^{W,m}(m_1(x_1),\ldots,m_1(x_n))$
3. $[\![DS]\!]_a{}^{W,m} \;=\; \lambda i \lambda t \lambda a.([\![D]\!]_a{}^{W,m})t([\![S]\!]_t{}^{W,m} i)a([\![S]\!]_a{}^{W,m} i)$
4. $[\![\hat{x}[F]]\!]_a{}^{W,m} \;=\; \lambda i.res(W,m)\downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid [\![F]\!]_a{}^{W,m[x/\langle X,i\rangle]}\}$
5. $[\![R_s]\!]_a{}^{W,m} \;=\; \lambda i.true$
6. $[\![S_1 \bigwedge S_2]\!]_a{}^{W,m} \;=\; \lambda i.([\![S_1]\!]_a{}^{W,m} i) \wedge ([\![S_2]\!]_a{}^{W,m} i)$
7. $[\![n(S)]\!]_a{}^{W,m} \;=\; \lambda i.[\![S]\!]_a{}^{W,m} i$
8. $[\![R]\!]_a{}^{W,m} \;=\; R_\perp{}'$
9. $[\![D]\!]_a{}^{W,m} \;=\; \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2.res(W,m)\Downarrow i \in Max_E(t_1 \cap t_2) \wedge a_1 \wedge a_2$
   if D is a distributive determiner
10. $[\![D]\!]_a{}^{W,m} \;=\; \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2.res(W,m)\Downarrow i \in Max(t_1 \cap t_2) \wedge a_1 \wedge a_2$ otherwise.

res applies restrictors to a DS on the basis of the map.

$res(W, m[x/\langle X,i\rangle]) = res(W \backslash \langle X,i\rangle, m)$

$res(W, [\,]) = W$

In clause 5 (for R_s), *true* just means that no restrictions are placed on the DS. The content will be set up by the corresponding truth conditional interpretation of R_s, appearing as t_1 or t_2 in the *Max* or Max_E term. The *Max* and Max_E part of the determiner clauses (9 and 10) ensure that the relevant DS component contains the right number of atoms and that they are of the right kind. The anaphoric conditions (a_1 and a_2) make sure that the DS slots for antecedents within the set terms are correct, and that the DS follows the right dependence relations. *Max* and Max_E will often yield a set containing only one member. The only time when this is not so is if the set it is applied to is not closed under summation, which can happen if it involves a number term. The definition of Max_E requires that the set terms are distributive; if they are not $Max_E(t_1 \cap t_2)$ is the empty set, and so no DS can satisfy the anaphoric conditions.

4.3.3.1 Examples

For (72), the anaphoric interpretation is (72.A), after making various logically valid simplifications.

(72) Most^i donkeys bray.

(72a) $(most\ donkey_s)(i)\ bray_s$

(72.A) $W \Downarrow i \in Max_E(donkey^* \cap bray^*)$

As discussed in section 4.2.5, example (72) does not structure the information in the DS. If instead of *bray*, a collective predicate like *gather* were used, then the formula would be unsatisfiable for any W, since $donkey^* \cap gather^*$ is not distributive, and hence $Max_E(\ldots)$ is the empty set.

For (73), we have

(73) Two[i] boys buy three[j] roses.

(73a) $(2_= \; boys_s)(i) \; \hat{x}[(3_= \; rose_s)(j) \; \hat{y}[buy(x,y)]]$

(73.A) $W \Downarrow i \in Max(boy^* \cap {}^*\{X \in \mathcal{D} \mid 3'_=(rose^*, buy^*(X))\}) \wedge$
$W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W \backslash (X,i) \Downarrow j \in Max(rose^* \cap buy^*(X)) \wedge$
$W \backslash (X,i) \downarrow j \subseteq \{Y \in \mathcal{D}_\perp \mid buy_\perp{}'(X,Y)\}\}$

This example can be used to explain in some detail how the interpretation of sentences with transitive verbs works. The principles also generalise to sentences containing ditransitive verbs and relative clauses. The first line just says that the content of slot i ($W \Downarrow i$) is a maximal collection of boys who buy three roses. The second line says that for each member X of slot i, the content corresponding to it in slot j ($W \backslash (X,i) \Downarrow j$) is all the roses bought by X. The final line says that the objects Y which make up this content stand in the buy relation to X, in the essential sense defined by buy'.

The only remaining thing to explain is why $\mathcal{D}_\perp$ rather than $\mathcal{D}$ and $buy_\perp{}'$ rather than buy' are used. The effect they have is as follows. Suppose that the sum of things in slot i is s. Because $s \oplus \perp = s$, $\perp$ can also appear in slot i. Substituting $\perp$ for X, we have $buy^*(X) = \emptyset$ because $\perp$ never appears in the extension of a predicate, and hence $Max(rose^* \cap buy^*(X)) = \{\perp\}$, giving $W \backslash (X,i) \Downarrow j = \perp$. This in turn means that $W \backslash (X,i) \downarrow j$ in the third line is $\{\perp\}$. Hence the use of $buy_\perp{}'$ to allow this case. It also follows that for $\perp$ in slot i, the only thing that can appear in slot j is $\perp$. The converse is not true: there can be an entity in slot i which corresponds to both $\perp$ and a non-$\perp$ member of slot j. Examples to show why this is needed are given below.

First, however, it is instructive to look at an example of how the formula works in a simpler case. Applying the model for (73) in section 4.3.2.5 to (73.A), we have that

$$Max(boy^* \cap {}^*\{X \in \mathcal{D}_\perp \mid 3'_=(rose^*, buy^*(X))\}) \;=\; Max(\{b_1 \oplus b_2\}) \;=\; \{b_1 \oplus b_2\}$$

and hence the content of slot i is $b_1 \oplus b_2$. This sum is equal to $\oplus\{\perp, b_1, b_2, b_1 \oplus b_2\}$, and each of these is therefore a possible member of $W \downarrow i$. The effect of substituting each in turn for X in the second line of (73.A′) is as follows:

1. $X = \perp$: $W \backslash (X,i) \Downarrow j = \perp$, as explained above

2. $X = b_1$: $W \backslash (X,i) \Downarrow j = r_1$.

3. $X = b_2$: $W \backslash (X,i) \Downarrow j = r_2 \oplus r_3$.

4. $X = b_1 \oplus b_2$: $W \backslash (X,i) \Downarrow j = r_1 \oplus r_2 \oplus r_3$

Finally, the condition in the third line allows the DS to contain $\langle \perp, \perp \rangle$, $\langle b_1, r_1 \rangle$ and $\langle b_2, r_2 \oplus r_3 \rangle$, the former from the definition of $buy_\perp{}'$, and the latter from the essential extension itself. The DS may not contain $\langle b_1 \oplus b_2, r_1 \oplus r_2 \oplus r_3 \rangle$, since it is not in the essential extension. This provides the correct structuring information.

By the reasoning given above, the following members may also appear in the DS: $\langle b_1, \perp \rangle$, $\langle b_2, \perp \rangle$ and $\langle b_1 \oplus b_2, \perp \rangle$. These are the alternative partitionings of the slot that may be needed for dependence relations with other antecedents, for examples like

(20) Three^i boys who buy five^j roses own four^k vases.

They also lead to the additional equivalent DSs described in section 4.2.5, which are harmless (if ugly). Sentences such as (20) provide part of the reason for using $\mathcal{D}_\perp$ and $buy_\perp{}'$. The other reason comes from the determiner *no*, when read intersectively. Substituting *no* for *two* in (73),

(78) No boys buy three roses.

has the effect that slot i contains $\perp$, and hence the only possible member of the DS is $\langle \perp, \perp \rangle$, which corresponds to the unacceptability of either anaphor in, for example,

(79) No^i farmer owns a^j donkey. ?They^i are cruel. ?They^i bray.

since there is nothing in either slot i or slot j for the pronouns to refer to. In the case of *no* in object position,

(80) Two^i boys buy no^j roses.

slot i may be non-empty, but the expression $W\backslash(X, i) \Downarrow j \in Max(rose^* \cap buy^*(X))$ forces slot j to be empty, since for any X in slot i, the argument of Max is the empty set: there are no roses bought by any of the boys. We will also need this mechanism when negation is introduced in section 4.5.4.

4.3.4 L(GQA) with anaphors

Anaphors are added to L(GQA) by means of two extra set terms. One draws on the whole of a specified slot of the DS, and one on the slot under restrictors. Both terms include a predicate P, called the agreement predicate, which can be *sg*, *pl* or *un*. The formation rules are:

F8a. If j is an index and P is a 1-place predicate, $Of(j, P)$ is a set term.

F8b. If j is an index and P is a 1-place predicate, $Of_r(j, P)$ is a set term.

Pronouns translate into quantifiers consisting of *the* applied to one of the set terms. Of_r is used where agreement under restriction is needed, and also in sentences where syntactic agreement has been made, such as

(81) Every farmer who owns a donkey beats it.

In the latter case, the agreement marker will be *un*, allowing both singular and plural entities. Examples (85) and (86) below show how this works. The procedure for deciding on which translation to use is discussed in section 4.4.1. Full NP anaphors are also translated using Of.

4.3.4.1 Interpretation

The truth conditional interpretation of Of is the input slot of the DS, allowing all possible sums formed from the slot. For Of_r, restrictors are applied.

$$[\![Of(j,P)]\!]_t^{W,m} = \lambda i.\natural(W\downarrow j)$$

$$[\![Of_r(j,P)]\!]_t^{W,m} = \lambda i.\natural(res(W,m)\downarrow j)$$

The anaphoric part just ensures that the whole of the output slot meets the agreement requirements. The result is the same for both cases of Of.

$$[\![Of(j,P)]\!]_a^{W,m} = \lambda i.P'(W\Downarrow i)$$

$$[\![Of_r(j,P)]\!]_a^{W,m} = \lambda i.P'(W\Downarrow i)$$

The expression $the'(\natural(W\downarrow i), Y)$ will often appear, and can be simplified to $W\Downarrow i \in Y$. In addition, agreement terms can be moved out of the set terms they appear in.

4.3.4.2 Examples

(82) and (83) are examples of sentences containing anaphors, with their translations and truth conditional and anaphoric interpretations. The subscripts on the NPs are the input indices.

(82) He_i^j whistles.

(82a) $(the\ Of(i,sg))(j)\ whistle_s$

(82.T) $the'(\natural(W\downarrow i), whistle^*)$

(82.A) $W\Downarrow j \in Max(\natural(W\downarrow i)\cap whistle^*) \wedge sg'(W\Downarrow j)$

Explanation: true if the whole of the input slot i whistles. The anaphoric part makes the output slot be the part of the input slot which whistles, i.e. all of it, and requires that the result be singular. The truth conditional part can be simplified, as discussed above, to

(82.T′) $W\Downarrow i \in whistle^*$

Essentially the same principles apply in the following example:

(83) They_i^k like them_j^l.

(83a) $(the\ Of(i,pl))(k)\ \hat{x}[(the\ Of(j,pl))(l)\ \hat{y}[like(x,y)]]$

(83.T) $W\Downarrow i \in lt$

(83.A)
$$\begin{aligned}&W\Downarrow k \in Max(\natural(W\downarrow i)\cap lt)\wedge\\ &W\downarrow k \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X,k)\Downarrow l \in Max(\natural(W\downarrow j)\cap like^*(X))\wedge\\ &\qquad\qquad W\backslash(X,k)\downarrow l \subseteq like_\perp{}'(X)\}\wedge\\ &pl'(W\Downarrow k)\wedge pl'(W\Downarrow l)\end{aligned}$$

where $lt = {}^*\{X \in \mathcal{D} \mid W \Downarrow j \in like^*(X)\}$
According to the truth conditions, slot i is a sum of things which like the whole of slot j. The anaphoric part sets slot k to be all of slot i which likes slot j, i.e. the whole of slot i. For each member X of slot k, the corresponding content of slot l is all that part of slot j which is liked by X. The output slots must both be plural and stand in the *like* relation.

To give an example, suppose that (83) follows (73) (*Two*^{i} *boys buy three*^{j} *roses*), and that we have the same model as before. In addition, suppose that each boy happens to like the rose or roses bought by the other boy. The extensions of *like* are then:

$$like' = \{\langle b_1, r_2\rangle, \langle b_1, r_3\rangle, \langle b_2, r_1\rangle\}$$

$$like^* = \{\langle b_1, r_2\rangle, \langle b_1, r_3\rangle, \langle b_2, r_1\rangle, \langle b_1, r_2 \oplus r_3\rangle, \langle b_1 \oplus b_2, r_1 \oplus r_2\rangle, \langle b_1 \oplus b_2, r_1 \oplus r_3\rangle, \langle b_1 \oplus b_2, r_1 \oplus r_2 \oplus r_3\rangle\}$$

The truth conditions are satisfied, since $W \Downarrow i$ is $b_1 \oplus b_2$, and

$$like^*(b_1 \oplus b_2) = \{r_1 \oplus r_2, r_1 \oplus r_3, r_1 \oplus r_2 \oplus r_3\}$$

of which $W \Downarrow j = r_1 \oplus r_2 \oplus r_3$ is a member. Working through the anaphoric part, the first line gives

$$W \Downarrow k \in Max(\{b_1, b_2, b_1 \oplus b_2\} \cap \{b_1 \oplus b_2\})$$

and hence slot k contains both boys. From the second line, the content of slot l corresponding to b_1 (if present in slot k) must be $r_2 \oplus r_3$, b_2 must correspond to r_1, and for $b_1 \oplus b_2$, slot l could contain $r_1 \oplus r_2 \oplus r_3$. Reasoning similar to that employed for (73) then means that we require the DS to contain $\langle b_1, r_2\rangle$, $\langle b_1, r_3\rangle$ and $\langle b_2, r_1\rangle$, and no other non-$\perp$ tuples are allowed. This DS also meets the agreement conditions. Thus a possible DS for (73) followed by (83) is (showing slots i, j, k and l):

$$\{\langle b_1, r_1, \perp, r_2\rangle, \langle b_1, r_1, \perp, r_3\rangle, \langle b_2, r_2 \oplus r_3, \perp, r_1\rangle\}$$

On this model, the more prominent reading is not available: a DS in which each boy likes the roses he bought cannot be constructed.

Example (84), on its complete full NP reading, has the translation and interpretation shown.

(84) The^j_i sheep graze.

(84a) $(the\ Of(i, pl))(j)\ (sheep_s \bigwedge graze_s)$

(84.T) $W \Downarrow i \in sheep^* \cap graze^*$

(84.A) $W \Downarrow j \in Max(\sharp(W \downarrow i) \cap sheep^* \cap graze^*) \wedge pl'(W \Downarrow j)$

The truth conditions force all of component j of the DS to be sheep which graze, and the anaphoric conditions require that component j is plural and has the same content as component i.

(85) illustrates bound anaphora, using the same indices for input and output. Note the translation of the pronoun using Of_r.

(85) Every^i cat washes itself^i_i.

(85a) $(every\ cat_s)(i)\ \hat{x}[(the\ Of_r(i, un))(i)\ \hat{y}[wash(x,y)]]$

(85.T) $every'(cat^*, wi)$

(85.A) $W \Downarrow i \in Max_E(cat^* \cap wi) \wedge$
$W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X,i) \Downarrow i \in Max(\sharp(W\backslash(X,i)\downarrow i) \cap wash^*(X)) \wedge$
$W\backslash(X,i)\downarrow i \subseteq wash_\perp{}'(X)\} \wedge$
$un'(W \Downarrow i)$

where $wi = {}^*\{X \in \mathcal{D} \mid W\backslash(X,i)\Downarrow i \in wash^*(X)\}$
Using the identity that $W\backslash(X,i)\Downarrow i = X$, wi simplifies to ${}^*\{X \in \mathcal{D} \mid X \in wash^*(X)\}$, and hence the truth conditions are that every cat is an X which washes X. The anaphoric part sets slot i to all such cats.

Finally, by translating the indefinite article as $1_{\geq}$, the universal reading of the quantified donkey sentence is obtained. Other readings are discussed in section 4.4.

(86) Every^i farmer who owns a^j donkey beats it^j_j.

(86a) $(every\ farmer_s\ \bigwedge\ \hat{x}[(1_{\geq}\ donkey_s)(j)\ \hat{y}[own(x,y)]])(i)$
$\hat{x}[(the\ Of_r(j, un))(j)\ \hat{y}[beat(x,y)]]$

(86.T) $every'(foad, bi)$

(86.A) $W \Downarrow i \in Max_E(foad \cap bi) \wedge$
$W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X,i) \Downarrow j \in Max(donkey^* \cap own^*(X)) \wedge$
$W\backslash(X,i)\downarrow j \subseteq own_\perp{}'(X)\} \wedge$
$W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X,i) \Downarrow j \in Max(\sharp(W\backslash(X,i)\downarrow j) \cap beat^*(X)) \wedge$
$W\backslash(X,i)\downarrow j \subseteq beat_\perp{}'(X)\} \wedge$
$un'(W \Downarrow j)$

where $foad = farmer^* \cap {}^*\{X \in \mathcal{D} \mid 1'_{\geq}(donkey^*, own^*(X))\}$
and $bi = {}^*\{X \in \mathcal{D} \mid W\backslash(X,i)\Downarrow j \in beat^*(X)\}$
Explanation: true if every farmer who owns one or more donkeys is such that $W\backslash(X,i)\Downarrow j$ is beaten by the farmer. This object is, from the anaphoric part, the donkeys owned by X. Slot i is all such farmers.

4.3.4.3 Acceptability of anaphors

The anaphoric part of the interpretation of (82), *he^j_i whistles* is

$$W \Downarrow j \in Max(\sharp(W \downarrow i) \cap whistle^*) \wedge sg'(W \Downarrow j)$$

Nothing else constrains what the output slot, j, is set to. Consequently, the first conjunct of the expression can always be satisfied; in the case where the input slot, i, does not whistle, the output slot will be empty, i.e. $W \Downarrow j = \perp$. The second conjunct, $sg'(W \Downarrow j)$, therefore determines whether the anaphoric part holds: it will fail to do so if either slot j is non-empty, but not singular, or if slot j is empty. The first case means that the anaphor is genuinely unacceptable. It would arise from, say,

(87) Some^i men are walking in the park. ?He^i whistles.

The second case would arise if the antecedent was singular, but did not whistle, as from

(88) Joe^i is walking in the park. He^i whistles.

in a case where Joe does not whistle. Ideally, we would like to be able to tell these apart. Examining the truth conditional part of the interpretation, $the'(\natural(W\downarrow i), whistle^*)$ shows that it is possible to do so. In the case where the antecedent does not whistle, then this formula is false, which provides sufficient information to distinguish the case in (87) from that in (88). Expressing the possibilities as a truth table:

Truth part	Anaphoric part	Meaning
False	False	sentence is false
False	True	
True	False	sentence is unacceptable
True	True	sentence is true

The unmarked case (the second line of the table) can never occur, since if the truth conditions are false, $W \Downarrow i$ does not whistle, and hence $W \Downarrow j$ is empty, in which case the singularity condition cannot be satisfied. Note also that the truth conditional part of the interpretation makes reference only to the input slot: $W \Downarrow j$ does not appear in it. Consequently, the body of the *Max* term in the anaphoric part (and hence the output slot) can be set up independently of the semantic number agreement test. Computationally, this might mean evaluating the truth conditions first, and if they hold, proceeding to the anaphoric conditions to determine acceptability. Similar principles can be extended to examples containing more than one anaphor. For further discussion of the issue of acceptability, see section 5.1.3.

Although it would be possible to develop this idea more formally, I intend to leave this as a subject for further work. The obvious approach is to define a three valued logic representing the three cases above, and to change the logical connectives (at least in the top-level interpretation) to capture the required result.

4.3.4.4 Proper nouns

In DRT, proper nouns receive a special treatment. It is Kamp's claim that (the discourse referents of) proper nouns are always accessible, in contrast to indefinite NPs. For example, (89) is acceptable, but (90) is not, in general.

(89) Every donkey likes John. He drives a tractor.

(90) Every donkey likes a farmer. ?He drives a tractor.

To ensure that proper nouns are always accessible, they are placed in the topmost DRS (Kamp, 1981, p.35).

In L(GQA), there is no structural notion of accessibility, and this approach cannot be followed. However, the interpretation achieves an equivalent result. In the interpretation of (89), each donkey will end up in a dependence relation with the extension of *john*, i.e. a single individual. *He* requires that the information it draws on is singular, which is clearly satisfied. In (90), there can be more than one farmer, and hence the singularity test fails. With plural proper nouns, the full range of dependence phenomena is possible:

(91) Some people know the Tebbits. They invite them to dinner.

(91) can mean that there is a group of people each of whom knows some of the Tebbits, and that each person invites the members of the Tebbits she or he knows to dinner. The semantics of *the* ensures that, in total, all of the Tebbits are known by the people in question.

4.3.5 Interpreting a discourse

A discourse is represented by a formula of the logic L(GQAD), which is an extended form of L(GQA). Two formation rules are added for conditionals and one for sequences of sentences. An alternative to the sequencing operator would be to add conjunction of formulae to L(GQAD). Except in rule D0, references to formulae mean formulae of L(GQAD). The syntax of L(GQAD) is as follows:

Basic expressions. As L(GQA) plus:

1. Conditional symbols: C. Examples: *usually*, *always*.

Derived expressions. As L(GQA).

Formation rules. As L(GQA) plus:

D0. If F is a closed formula of L(GQA), then F is a formula of L(GQAD).

D1. If F_1 and F_2 are formulae, then $F_1; F_2$ is a formula.

D2. If F_1 and F_2 are formulae, and C is a conditional symbol, then $C(F_1, F_2)$ is a formula.

D3. If F_1 and F_2 are formulae, C is a conditional symbol, and i is an index, then $C(i)(F_1, F_2)$ is a formula.

Rule D1 forms sequences of sentences. Rules D2 and D3 are for conditionals and sentences which receive a similar interpretation, such as:

(92) Whenever John kisses a frog, it turns into a toad.

D2 covers symmetric readings and D3 ones asymmetric on the antecedent with index i.

The interpretation function for L(GQAD) is I, and takes an L(GQAD) formula and a set of DSs to a set of DSs. The basic rule is:

$$I(F, V) \;=\; \{W \in V \mid [\![F]\!]^W\}$$

where V is a set of DSs and $[\![.]\!]$ is the conjunction of the truth and anaphoric interpretation functions, as already defined. Interpretation of discourse starts with the set of all possible DSs. The rule can be seen as discarding all those DSs that are inconsistent with the constraints imposed by F. A discourse as a whole can be considered true if the set of possible DSs is non-empty, and false otherwise.

There are several possible definitions for sequencing: first to intersect the set of possible DSs for each formula, i.e.

$$I(F_1; F_2, V) \;=\; I(F_1, V) \cap I(F_2, V)$$

Alternatively, we could take a more dynamic view, where V is filtered through F_1 and the result filtered through F_2:

$$I(F_1; F_2, V) \;=\; I(F_2, I(F_1, V))$$

It is relatively unimportant which definition is used. A disadvantage of this approach is that if any sentence is false, any sentences after the false one are interpreted with respect to an empty set of DSs, and hence will be judged false as well. This is clearly not what actually happens in discourse processing. In

(93) John owns a car. Bill owns a bike.

the second sentence is interpretable even if the first is false. One way in which this might be rectified is to redefine the interpretation of $F_1; F_2$, so that the set of DSs obtained by interpreting F_1 is passed to F_2 if non-empty, and otherwise F_2 is passed the DSs in force prior to F_1. A suitable definition is

$$I(F_1; F_2, V) = \begin{cases} I(F_1, V) \cap I(F_2, V) & \text{if } I(F_1, V) \neq \emptyset \\ I(F_2, V) & \text{if } I(F_1, V) = \emptyset \end{cases}$$

The interpretation for rule D2 is:[5]

$$I(C(F_1, F_2), V) \;=\; \{W \in V \mid C'(I(F_1, V), I(F_1, V) \cap I(F_2, V))\}$$

The formula passes on either the whole of V, unchanged, or none of it depending on whether the conditional is true or not. The definitions of C', the meta-language counterpart of C, exploit the similarity noted between adverbs of quantification and determiners noted by Stump (1984), expressing them as relation between the cardinalities of $I(F_1, V)$ and of $I(F_1, V) \cap I(F_2, V)$, just as p and $p \cap q$ are used for generalized quantifiers. Some suitable definitions are:

always: $\lambda mn.m = n$ (same as *every*).

usually: $\lambda mn.m \geq n/2$ (same as *most*).

rarely: $\lambda mn.n \leq r(m)$, for some measure of "rareness" r (compare *few*).

sometimes: $\lambda mn.n \geq 1$ (same as *some*).

never: $\lambda mn.m = 0$ (same as *no*).

According to these definitions, the structure as well as the content of the slots is relevant, since both distinguish one DS from another. I think this is not quite right. My intuition about (94),

(94) If two boys buy three roses, they usually like them.

is that we only need to know how the boys and roses are partitioned for checking whether the *buy* and *like* relations produce the same partitioning. It may be necessary to consider structuring information where modal operators are used to abstract away from the question of whether the same objects can be bought by more than one individual:

(95) If two boys can afford to buy three roses, they will usually like them.

If there are different ways that the same boys can get together to buy the same collection of roses, then they could be taken as different cases for testing whether *usually* holds. However, this example can just as well be explained by saying that *usually* is quantifying over (something like) situations: that is, it is an instance of the phenomenon we saw in section 2.3.2, for the sentence

(76) If a man has the same name as another man, he usually avoids addressing him by name.

[5]The formulation will be slightly revised below.

The interpretation rule for D1 is therefore revised to ignore the structure of DSs for the purpose of counting how many of them there are:

$$I(C(F_1, F_2), V) \;=\; \{W \in V \mid C'(\sum I(F_1, V), \sum I(F_2, I(F_1, V)))\}$$
$$\text{where } \sum V = \{\oplus W \mid W \in V\}$$

For example, if the following are both members of V:

$$\{\langle a_1, b_1\rangle, \langle a_2, b_2\rangle, \langle a_3, b_3\rangle\}$$

$$\{\langle a_1, b_1\rangle, \langle a_2 \oplus a_3, b_2 \oplus b_3\rangle\}$$

then only the following appears in $\sum V$:

$$\{\langle a_1 \oplus a_2 \oplus a_3, b_1 \oplus b_2 \oplus b_3\rangle\}$$

since the two DSs are distinguished only by how they structure the information.

Antecedents within the conditional are rendered unavailable for further reference. In general, this seems to be correct. The second sentence of

(96) If a farmer owns a donkey, he beats it. They are cruel to them.

is not acceptable, except possibly on a generic interpretation, meaning that farmers in general are cruel to donkeys in general. Sometimes it does seem to be possible to export antecedents, as in

(97) If Bill/a (specific) farmer owns a donkey, he beats it. He is cruel.

In section 4.5.2, I will discuss a means of dealing with such examples.

For asymmetric readings, rule D3 is used, one discourse referent being singled out by the index i. The same quantifiers for the adverb symbols are used, but they now quantify over DSs which are distinguished from one another purely on the basis of the contents of slot i. Thus for the reading of (98) which is asymmetric on slot i,

(98) If three^i boys buy five roses, they^i usually like them. (Whereas if four girls buy five roses, they usually don't.)

it does not matter if one of the possible groups of three boys in fact bought two different lots of five roses. This is easily obtained by discarding all of the DS other than slot i. We first need to extend the definition of extraction ($\Downarrow$) to sets of DSs:

$$V{\downarrow}i = \{W{\downarrow}i \mid W \in V\}$$

The interpretation for rule D2 is then:

$$I(C(i)(F_1, F_2), V) \;=\; C'(\sum I(F_1, V){\downarrow}i, \sum I(F_2, I(F_1, V)){\downarrow}i)$$

As with symmetric conditionals, it the the content and not the structure of the DS that matters, and hence $\sum$ is used.

4.4 Translating from English to L(GQA)

The DS model and the interpretation of L(GQA)/L(GQAD) each capture part of the empirical content of TAI. The remainder is provided by the procedure for translating from English sentences to L(GQA)/L(GQAD) formulae. In particular, the translations of determiners and pronouns lead to further discussion of some issues raised in chapter 2: readings of donkey sentences, and structural versus semantic constraints on the antecedent-anaphor relation.

There are a number of places where more than one translation is possible. Preference rules are suggested to select one translation as the better one. The idea behind such rules is that they are applied unless either the result is false (or unacceptable) in which case a different translation must be used, or unless there is some additional information which selects a dispreferred translation. For the time being I leave this mechanism fairly vague. It is discussed further in section 5.2.1.

4.4.1 Sentence grammar

The grammar covers a small fragment of English, using a context free syntax. The notation $C : s$ is used to indicate a node with syntactic category C and semantic translation s. Semantic translations are combined using lambda reduction, with the notation fx meaning f applied to x. There is nothing to enforce subject-verb or determiner-nominal agreement in the grammar, or to put pronouns in the right case. It would be a straightforward exercise to modify the syntax rules for them, for example by using a feature-based unification grammar (see Shieber (1986) for an introduction).

Syntactic categories

1. S (sentence)
2. NP (noun phrase), N′ (nominal), N (noun), PN (proper noun), Pro (pronoun)
3. VP (verb phrase), IV (intransitive verb), TV (transitive verb), DV (ditransitive verb)
4. Rel (relative pronoun)
5. Num (numeral)

Production rules

S1. S:qs $\rightarrow$ NP:q VP:s

S2. NP:$p(i)$ $\rightarrow$ PN:p

S3. NP:$p(i)$ $\rightarrow$ Pro:p

S4. NP:$(dp)(i)$ $\rightarrow$ Det:d N′:p

S5. NP:$\lambda s_1.q(s_2 \bigwedge s_1)$ $\rightarrow$ NP:q Rel VP:s_2

S6. VP:p_s $\rightarrow$ IV:p

S7. VP:$\hat{x}[q\ \hat{y}[p(x,y)]]$ $\rightarrow$ TV:p NP:q

S8. VP:$\hat{x}[q_1\ \hat{y}[q_2\ \hat{z}[p(x,y,z)]]]$ $\rightarrow$ DV:p NP:q_1 NP:q_2

S9. N′:$s_1 \bigwedge s_2$ $\rightarrow$ N′:s_1 Rel VP:s_2

S10. N′:p_s $\rightarrow$ N:p

S11. NP:$(\Phi\ n(p)) \rightarrow$ Num:n N$'$:p

Rule S5 is for non-restrictive relative clauses and rule S9 for restrictive relatives. The procedure for assigning indices i and j in S2–S4 is discussed below.

A sample lexicon is as follows:

1. Nouns: *farmer*, *donkey*, *man*, *woman*. Semantic translation: 1-place predicate symbols.
2. Proper nouns: *john*, *mary*. Semantic translation: $(\textit{the}\ n)$, where n is a 1-place predicate symbol.
3. Intransitive verbs: *walk*, *gather*. Semantic translation: 1-place predicate symbols.
4. Transitive verbs: *own*, *buy*, *beat*. Semantic translation: 2-place predicate symbols.
5. Ditransitive verbs: *give*. Semantic translation: 3-place predicate symbols.
6. Relative pronouns: *who*, *which*.
7. Determiners: *every*, *some*, *few*, *most*, *no*. Semantic translation: L(GQA) determiner. Also: *a*, *the*, *one*, *two*, *three*, etc., which receive special treatment detailed below.
8. Pronouns: *he*, *him*, *she*, *her*, *it*, *they*, *them*. Semantic translation: $(\textit{the}\ \textit{Of}(j,P))$, or $(\textit{the}\ \textit{Of}_r(j,P))$; see below.

Numerals. Numeral determiners may be translated as: $n_=$ (exactly), $n_\geq$ (intersective, at least) and $n_\leq$ (at most). In addition, the translation $\lambda s.(\Phi\ n(s))$ gives the specific, at least reading.

Pronouns. In the pronoun translations, j is the input index, and P is the agreement predicate. The agreement predicate for *he*, *she*, *it*, *him* and *her* is *sg*, and for *they* and *them*, *pl*. No distinction is made for gender. It would be a straightforward matter to add further agreement predicates for this. The default translation is the one which uses *Of* rather than $\textit{Of}_r$. Other translations are sometimes needed: see section 4.4.4 below.

"The". When used for complete full NP anaphors, the definite determiner *the* is translated as

$$\lambda s_1.\lambda i.\lambda s_2.(\textit{the}\ \textit{Of}(j,P))(i)\ (s_1 \bigwedge s_2)$$

where j is the input index. P is *sg* for singular noun phrases and *pl* for plural ones.

Non-anaphoric *the* can be translated in exactly the same way, with the same values for the input and output indices. From a sentence such as

(99) The man walks.

the translation is

(99a) $(\textit{the}\ \textit{Of}(i,sg))(i)\ \textit{walk}_s$

the two parts of the interpretation are

(99.T) $\textit{the}'(\sharp(W\downarrow i), \textit{man}^* \cap \textit{walk}^*)$

(99.A) $W \Downarrow i \in \textit{man}^* \cap \textit{walk}^* \wedge \textit{sg}'(W \Downarrow i)$

which require the i slot contains a single man that walks.

4.4.2 Indices

It is straightforward to assign indices to non-anaphoric NPs: just use a new index for each such NP, perhaps by assigning the ith index to the ith NP that occurs in the discourse. Anaphoric NPs have input and output indices. The input index can either be assigned the same value as that of the antecedent, or it can be assigned a new index, and a condition imposed on DSs that they contain the same information in the anaphor and antecedent slots. In L(GQA) terms, this means that $W \downarrow i = W \downarrow j$, where i and j are the indices involved. In either case, the possible antecedents have first to be identified. This is a complex problem in itself, and I will not address it here; see section 2.4 for references, mainly from the computational literature. For a given occurrence of an anaphor, it may not be possible to pick a single candidate antecedent, or to decide on the best of the available possibilities, until more of the discourse has been processed, as (100) and (101) illustrate.

(100) John saw Bill in the park. He said "Hello" and John hit him.

(101) John saw Bill in the park. He said "Hello" and Bill hit him.

The antecedent of *he* cannot be definitely established on first reaching it. After processing the final conjunct, it becomes more likely that it is Bill in the first case and John in the second.

For anaphoric NPs, we also have to choose whether the input and output indices should be the same. Giving them different values means that the content of the two slots will be the same, but the anaphors may structure the information differently from their antecedents, establishing a different dependence relation. Data in section 4.2 showed that in sentences with two or more anaphors to antecedents that stand in a dependence relation, a reading where the same dependence relation is followed by the anaphors is often prominent. To capture this reading in L(GQA), the same index is used for input and output. The cost of this approach is that it increases the amount of ambiguity: there are two possible translations per anaphor. The prominent reading suggests the following preference rule:

Preference rule 1 (indices) Assign anaphor translations the same input and output indices by preference.

4.4.3 Discourse grammar

The grammar so far covers single, non-conditional sentences. To complete the picture, we need to extend the grammar to cover discourses and conditionals. Translations are into L(GQAD).

Syntactic categories. As above, plus

1. D (discourse).
2. Adv (adverb of quantification).

Production rules. As above, plus

S12. D:f $\rightarrow$ S:f

S13. D:$f_1; f_2$ $\rightarrow$ D:f_1 S:f_2

S14. D:$c(f_1, f_2)$ $\rightarrow$ Adv:c S:f_1 S:f_2

S15. D:$c(i)(f_1, f_2)$ $\rightarrow$ Adv:c S:f_1 S:f_2

Rule S12 starts off a discourse from a single sentence, and rule S13 extends the discourse. Rule S14 covers symmetric conditionals, and rule S15 asymmetric ones. In each case i is an index appearing in a quantifier of the left hand sentence. The category of adverbs (a poor approximation to the syntactic use of the term) includes *if*, *usually if*, *always if*, *whenever* and so on. Adverbs are translated as L(GQAD) conditional symbols, with bare *if* translating as *always*. The grammar is rather simplistic as it stands. It will reject some conditional sentences on syntactic grounds, which by rights should be allowed, such as

(102) If a farmer owns a donkey, he usually beats it.

(103) If a farmer owns a donkey and an innkeeper owns a stable, the farmer usually keeps the donkey in the stable.

(104) A farmer, if he owns owns a donkey, beats it.

However, the syntax rules as they stand suffice to illustrate the semantic approach taken to conditionals.

4.4.4 Determiner translations and agreement

The most natural translation of *a* is the specific, at least reading, $\lambda s.(\Phi\ 1(s))$. On this translation the sentence

(105) A man walks in the park. He whistles.

can be true if there is more than one man walking in the park, but one such man is singled out for subsequent reference. By analogy with the numerals, we might ask whether any of the other translations are possible. There seems to be little or no argument for $1_{\leq}$, i.e. at most one. The exactly reading, $1_{=}$, is sometimes possible, particularly under suitable intonation:

(106) This town doesn't have THREE churches, it has A church.

In simple examples, the non-specific, at least translation ($1_{\geq}$) has little going for it. It would permit the second sentence of (107).

(107) A cat is digging up my rose bush. ?They are burying frogs.

The only case where *they* seems at all acceptable is on a reading where it is taken as generic or as referring to contextually given collection, as when (107) is followed by:

(108) They baffle me when they do that.

Here, *they* could be cats in general, or perhaps all cats in the neighbourhood, but certainly not the intersective collection of cats that are digging up my rose bush (right now). Link (1987) makes a similar point, quoting as an example

(109) A third grader solves such a problem easily.

following which *they are clever enough* would be OK. (105) and (107) are also consistent with the agreement predicates proposed for pronoun translations.

In some cases, such as donkey sentences on a universal reading, *a* must receive a different translation. To obtain these readings, I define a *variant translation principle* (VTP), which acts during or after the translation process. The idea of the principle is to use properties of other determiners in the sentence, and the relation in which they stand to *a* to predict which translation to use. As the discussion in chapter 2 revealed, the principle must be treated with caution. There is considerable variation in what is judged to be the correct reading, depending on the lexical items, context, speaker, stress and so on. The VTP must apply after quantifier scoping, and so is best stated in terms of L(GQA) expressions.

In an expression $(D\ S_1)(i)\ S_2$, the set term S_1 is called the restriction and S_2 the body of D. The lexical entry of every determiner specifies two variant translations for *a*, one to be used when *a* appears in the restriction and one when it is in the body of the determiner that triggers the VTP. The variant need not be the same for both cases.

Variant translation principle (VTP).

> If a NP whose determiner is *a* is translated as a quantifier occurring in the restriction or body of a determiner D, then it may be translated using the appropriate variant specified in the lexical entry for D.

If the VTP applies, D is said to trigger it. To see how the principle works, consider (110)–(113).

(110) Every farmer who owns a donkey beats it.

(111) Every farmer owns a donkey. They beat them.

(112) Most farmers who own a donkey beats it.

(113) Most farmers own a donkey. They beat them.

The universal reading of (110) follows if the restrictor variant translation specified for *every* is $1_{\geq}$. This is one interpretation of the data; if the view of Kadmon (1990) is accepted instead, that the quantified donkey sentence says nothing about farmers who own more than one donkey, then the variant translation should be $1_{=}$. Both variants are also possible for *most*, although the discussion of (112) in section 2.3.1 perhaps suggests that $1_{=}$ is the better one. For both *every* and *most*, the body variant translation seems, on the evidence of (111) and (113), to be $1_{\geq}$. Note that the change in translation from $\lambda s.(\Phi\ 1(s))$ to $1_{\geq}$ is a change in the anaphoric properties alone, and not the truth conditions.

The variant translation is not always correct, as the data of section 2.3 showed. However, the widespread acceptance of the universal reading for donkey sentences suggests that it is probably to be preferred:

Preference rule 2 (variant translations) Apply the variant translation principle by preference.

It is also interesting to question whether the VTP ever fails to apply to indefinites in the body, as in (111) and (113). I think that variant translation is much more strongly preferred in these cases than it is when the indefinite is in the restriction, as in (110) and (112). However, it is possible to find contexts where the specific, at least reading is better. An example is this: suppose that a novice computer user goes into a store, and asks the salesman how he can learn to use whichever brand he buys. The salesman replies

(114) Every computer comes with a manual. They say how to get started.

They can refer to a specific manual per (brand of) computer, but (114) should not rule out the possibility of computers having more than one manual, say an introduction and a reference guide. The specific, at least translation of a, $\lambda s.(\Phi\ 1(s))$ captures exactly this.

To complete the account, we also need to specify when non-standard pronoun translations are to be used. The special cases of partitives (*one of them*) and floating quantifiers (*they each beat it*) are omitted for the time being; see section 5.3.2. The cases we are concerned with were exemplified by (95)–(99) of section 2.4, and (85)–(86) above, namely:

(85) Every cat washes itself.

(86) Every farmer who owns a donkey beats it.

Here, the pronoun must be translated as the restricted form, with an unspecified agreement predicate; that is, the translation is $(the\ Of_r(i, un))(j)$, where i is the antecedent's index. A syntactic characterisation was suggested in section 2.4. However, there is a reformulation in terms of L(GQA) formulae which is more concise.[6] The principle can be stated as:

Pronoun Variant Translation Principle (PVTP).
A pronoun is translated as $(the\ Of_r(i, un))(j)$, if:

The antecedent translates as a quantifier Q_1 and the pronoun as a quantifier Q_2 such that

EITHER Q_1 combines with a set term S containing Q_2

OR there is a quantifier Q which immediately encloses Q_1 and which combines with a set term containing Q_2.

When the conditions are satisfied, the antecedent and pronoun must agree syntactically.

The first case applies to (85), and the second to (86). Q "immediately encloses" Q_1 if Q consists of a determiner applied to a set term $\hat{x}[Q_1 \ldots]$ or to a conjunction of set terms one of which has this form. A quantifier Q combines with a set term S if they appear in a formula of the form $Q(i)\ S$. An example where the pronoun is *not* to be given the variant translation is

(115) Most shepherds who know every farmer who owns a donkey beat ?it.

In this case, the quantifier which immediately encloses *a donkey* is derived from *every farmer ...*, while it is the one derived from *most shepherds ...* which combines with the set term containing the anaphor, and hence the PVTP does not apply. Different input and output indices are allowed in the pronoun translation. An example where they might be needed is

(116) Every farmer who owns two^i oxen harnesses them$_i^j$ to a plough.

[6] For some related discussion, see Bosch (1983), who likewise considers syntactic constraints in anaphora (proposed by Reinhart (1983)) from the point of view of functor-argument structure.

The farmer can own each ox individually, but only be able to harness them to the plough together, so that slots i and j have the same content but different structure.

As with the VTP, expressing the PVTP in terms of the semantic translation has the consequence that it can be applied after quantifier scoping. This appears to be correct. Taking (115) as an example, if *a donkey* is given widest scope, then the anaphor becomes acceptable. One way in which the PVTP is not entirely satisfactory is that it introduces structural notions into the theory. There is an interesting contrast with DRT, where much of the theory is based on structural relations supported by a small amount of semantic information. TAI draws principally on semantic information, with some support from structural properties.

In the examples examined so far, the determiners have forced a distributive reading. It may be necessary to make the PVTP optional with other determiners. For example, there is a reading of

(117) Three boys who bought five vases broke them in a fight.

in which the boy who broke any particular vase was not necessarily the one who bought it. In this case, the standard pronoun translation is required.

The logic as it stands can generate three of the four possible readings of donkey sentences, depending on the translation of a: $1_{\geq}$ for the universal reading, $\lambda s.(\Phi\ 1(s))$ for the unique anaphor one, and $1_{=}$ for the unique antecedent reading. There is nothing at present for the indefinite lazy reading. One way of adding it would be to translate the anaphor using *some* as the determiner rather than *the*, that is as $(some\ Of_{r}(i, P))(j)$. The interpretation will then require that only part of slot i is referred to by the pronoun, and this part of it is placed in slot j. The conditions for when the reading is available are less clear. I personally find it to be the least preferred reading in general, and while it would be possible to state a preference rule which indicates this, there seems to be little "empirical content" in doing so. The preference for other readings probably comes from more general pragmatic principles. Taking the standard quantified donkeys sentence *every farmer who owns a donkey beats it* as an example, the universal reading and the unique anaphor and antecedent readings allow the hearer to identify precisely what the referent of *it* is for each farmer. This is not so for the indefinite lazy reading: for a farmer who owns more than one donkey, it says that the farmer beats some of the donkeys. The hearer must therefore either identify, from other information or randomly, which collection of donkeys is meant, or must accept that the exact referent of the pronoun cannot be established. A possible way of identifying when the indefinite lazy reading is available might therefore be to test whether these conditions on the hearer can be reasonably assumed. I will return to the discussion of such pragmatic factors in section 5.2.1.

One way in which my approach differs from Kadmon's is that she makes the translation of a change only if there is a subsequent anaphor. I predict the change from the other NPs in the sentence where it occurs, and make the pronoun translation change in response to the antecedent. Her theory is rather unsatisfactory from a point of view of left-to-right processing of a discourse, in that the interpretation of the first sentence changes from universal to unique when the second sentence is processed, and is then once again made universal when the various things the antecedent could refer to are indistinguishable. These problems do not occur with the principle I propose.

4.4.4.1 Conditionals

For conditional donkey sentences, exactly the same variant translation principles as above are employed. The translation of

(118) If a farmer owns a donkey, he beats it.

will come out as

$$\text{(118a)}\quad always((\Phi\ 1(farmer_s))(i)\ \hat{x}[(\Phi\ 1(donkey_s))(j)\ \hat{y}[own(x,y)]], \\ (the\ (Of(i,sg))(i)\ \hat{x}[(the\ (Of(j,sg)))(j)\ \hat{y}[beat(x,y)]])$$

using the natural translations of *a* and pronouns arrived at above. The formula gives the symmetric reading: it takes every DS containing one farmer in the i slot and one donkey he owns in the j slot, and checks that the slots also stand in the beat relation. If any farmer owns more than one donkey, there will be a separate DS for each donkey he owns. Note that the output indices of the pronouns are the same as their input indices. There is in fact no need to do this; since the DS slots contain a single thing the only possible restructuring of the information is the same as the original structure. In (119)

(119) If two farmers each own a donkey, they beat them.

restructuring is possible, for example in a situation where each of a pair of farmers maliciously beats the donkeys belonging to the other farmer.

Example (119) reveals another point as well. Suppose farmer f_1 owns donkeys d_1 and d_2, f_3 owns d_3 and f_4 owns d_4. Because *a* translates as $\lambda s.(\Phi\ 1(s))$ and hence picks out one donkey for each farmer, there are five possible DSs:

$\{\langle f_1, d_1\rangle, \langle f_3, d_3\rangle\}$

$\{\langle f_1, d_1\rangle, \langle f_4, d_4\rangle\}$

$\{\langle f_1, d_2\rangle, \langle f_3, d_3\rangle\}$

$\{\langle f_1, d_2\rangle, \langle f_4, d_4\rangle\}$

$\{\langle f_3, d_3\rangle, \langle f_4, d_4\rangle\}$

Interpreting *if* as *always*, this causes no problems. If there is an explicit adverb such as *usually*, the truth conditions will be based on the proportion of the above DSs which satisfy the consequent sentence, on a symmetric reading. However, TAI predicts another possible reading of the sentence. Besides the readings asymmetric on the farmers or the donkeys, there may also be one where DSs are not distinguished on the basis of the two donkeys belonging to f_1, so that only three DSs are constructed:

$\{\langle f_1, d_1\rangle, \langle f_1, d_2\rangle, \langle f_3, d_3\rangle\}$

$\{\langle f_1, d_1\rangle, \langle f_1, d_2\rangle, \langle f_4, d_4\rangle\}$

$\{\langle f_3, d_3\rangle, \langle f_4, d_4\rangle\}$

In L(GQA) terms, the reading is obtained by allowing the variant translation principle to apply to *a donkey* giving it a determiner of $1_{\geq}$, triggered by the presence of *two*. Furthermore, the different translations do have different truth conditions. Suppose that the adverb relating the antecedent and consequent sentences is *usually*, and that all farmers beat their donkeys, with the one exception that f_1 does not beat f_2. Then the first set of DSs gives a true result, since there are three out of the five DSs which satisfy the consequent. However, the second translation comes out false: only the last of the three DSs holds in the consequent. That is, if the variant translation principle is allowed to apply here, TAI predicts that there are two readings which are asymmetric on the farmer slot, rather than just one. I find myself unable to arrive at any definite conclusion about how the readings should be narrowed down or placed in order. It critically rests on how we decide to define cases in the sense of section 2.3.2. Again, some suggestions are discussed in section 5.2.1.

4.4.5 Modification rules

There are two rules which manipulate L(GQA) formulae to produce alternative quantifier scopings.

Quantifier scope rules.

1. A formula of the form
   $Q_1(i)\ \hat{x}[Q_2(j)\ \hat{y}[F]]$
   may be replaced by
   $Q_2(j)\ \hat{y}[Q_1(i)\ \hat{x}[F]]$
   provided there are no free occurrences of x in Q_2 or of y in Q_1.
2. A formula of the form
   $(D_1\ (S_1 \wedge\ \hat{x}[Q(j)\ \hat{y}[F]]))(i)\ S_2$
   may be replaced by
   $Q(j)\ \hat{y}[(D_1(S_1 \wedge\ \hat{x}[F]))(i)\ S_2]$
   provided there are no free occurrences of x in Q or of y in Q_1.

The first rule allows the direct translation of

(120) Every^i child reads some^j books.

to be changed to one where there are some books read by every child, i.e. from (120a) to (120b).

(120a) $(\mathit{every\ child})(i)\ \hat{x}[(\mathit{some\ book})(j)\ \hat{y}[\mathit{read}(x, y)]]$

(120b) $(\mathit{some\ book})(j)\ \hat{y}[(\mathit{every\ child})(i)\ \hat{x}[\mathit{read}(x, y)]]$

The second rule will change the direct translation of

(121) Two^i boys who buy three^j roses walk.

from (121a) to (121b):

(121a) $(2_{=}\ (\mathit{boy} \wedge\ \hat{x}[(3_{=}\ \mathit{rose})(j)\ \hat{y}[\mathit{buy}(x, y)]]]))(i)\ \mathit{walk}$

(121b) $(3_{=}\ \mathit{rose})(j)\ \hat{y}[(2_{=}\ (\mathit{boy} \wedge\ \hat{x}[\mathit{buy}(x, y)]]))(i)\ \mathit{walk}]$

The modified translation can be paraphrased as: there are three roses which are such that two boys who buy them walk.

Quantifier scoping can be done another way, using the DS. To give a quantifier wide scope, we move it from its original position, and in its place leave an *Of* expression for the same slot of the DS.

Widening rule.

Given a formula F containing a quantifier Q with index j, F may be replaced by $Q(j)\ \hat{u}[F']$, where u is a variable that does not occur in F and F' is obtained from F by replacing Q with $(the\ Of_r(j, un))$. The rule may only apply if no variables in Q are free in their new positions.

Intuitively the rule can be seen as paraphrasing the object wide scope reading of (120) as:

(122) There are some books such that every child reads them.

Applying the widening rule to (120) yields:

(122a) $(some\ books)(j)\ \hat{u}[(every\ child)(i)\ \hat{x}[(the\ Of_r(j, un))(j)\ \hat{y}[read(x, y)]]]$

which has interpretation:

(122.T) $some'(book^*, cr)$

$$\begin{aligned}
\text{(122.A)}\ & W{\Downarrow}j \in Max_E(book^* \cap cr) \wedge \\
& W{\downarrow}j \subseteq \{U \in \mathcal{D}_\perp \mid W\backslash(U,j){\Downarrow}i \in Max_E(child^* \cap tr(U)) \wedge \\
& \qquad W\backslash(U,j){\downarrow}i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(U,j)\backslash(X,i){\Downarrow}j \in \\
& \qquad\qquad Max(\sharp(W\backslash(U,j)\backslash(X,i){\downarrow}j) \cap read^*(X)) \wedge \\
& \qquad\qquad W\backslash(U,j)\backslash(X,i){\downarrow}j \subseteq read_\perp{}'(X)\}\} \wedge \\
& un'(W{\Downarrow}j)
\end{aligned}$$

where $cr\ =\ {}^*\{U \in \mathcal{D} \mid every'(child^*, {}^*tr(U))\}$
and $tr(U)\ =\ \{X \in \mathcal{D} \mid the'(\sharp(W\backslash(U,i)\backslash(X,j){\downarrow}j), read^*(X))\}$
Noting that $W\backslash(U,j)\backslash(X,i){\Downarrow}j = U$, the truth-conditional part of this says that there are some books which are U's such that every child reads U. From the anaphoric part, slot j is the collection of all such books.

There is also a third possibility, which I will not explore in detail, and this is to treat quantifier scoping as a constraint applied to the DS. For example, to make *some books* with index j have wide scope in *every*^{i} *child reads some*^{j} *books*, we want each part of slot j corresponding to a member of slot i to be the same, so that every child in slot i reads the same books in slot j. Since this also means that collection of books per child is the same as the collection of books in total, the constraint on the DS can be stated as:

$$\forall X \in W{\downarrow}i[W\backslash(X,i){\Downarrow}j \neq \perp \rightarrow W\backslash(X,i){\Downarrow}j = W{\Downarrow}j]$$

An advantage of this approach is that the decision about quantifier scope need not be made when the sentence is processed: the scoping constraints can be added at any stage.

Some preference rules for quantifier scoping are suggested by, for example, Moran and Pereira (1992); again, see section 5.2.1 for further discussion.

The modification rules appear to violate the principle of compositionality, in that there is no longer a one-to-one correspondence between the syntactic structure of a constituent and its L(GQA) translation. The same is true of many analyses of quantifier scope.[7] One way of regaining compositionality, albeit not an entirely satisfactory one, is to find all of the syntactic rules and combinations of syntactic rules which lead to the the formulae triggering the modification rules. An extra collection of syntax rules can then be constructed which

[7]Although see Kempson and Cormack (1981) for an attempt to rectify this

take the same lexical items to a different configuration, corresponding to the modified translation. The cost of doing so is an explosion in the number of syntactic rules (and a very odd syntactic theory). We could similarly trace through the variants to the interpretation functions, and build the ambiguity into the logic. This perhaps indicates that it is more important to follow the spirit of compositionality, of being able to identify where all the elements of the translation come from, than a more rigid statement of it.

4.5 Extensions to the theory

Two extensions to the theory are examined in this section: non-intersective determiners, which were introduced in section 4.2.4.1, and logical connectives. Most formal theories of semantics try to establish a close correspondence between the logical connectives of negation, conjunction and disjunction, and natural language terms such as *not*, *and* and *or*. I think it is generally misguided to do so, just as the semantics of logical implication are impoverished with respect to those of *if*. In this section, I will give a brief review of some of the data on the natural language terms, and propose suitable logical connectives for L(GQA). The account is by no means complete, particularly as regards negation. Kamp and Reyle (1990) has been a valuable source for the data in this section.

4.5.1 Non-intersective determiners

The subject of non-intersective determiners was raised in section 4.2.4.1. To recap, intersective determiners pass on the intersection of the nominal extension and the predicate to which they apply to anaphors, whereas non-intersective ones pass on simply the extension of the nominal. For example, in (123), *most* is intersective, while in (124), *few* behaves non-intersectively.

(123) Most linguists smoke, although they (= linguists who smoke) know it causes cancer.

(124) Few linguists smoke, since they (= all linguists) know it causes cancer.

The terms originate with Webber (1979), from whom these examples are also taken.

L(GQA) as presented so far, makes all determiners intersective, but it is easy to modify the semantics to allow non-intersective uses. L(GQA) determiner symbols are marked as either intersective or non-intersective (just as some determiners were marked as distributive before) and the following extra rules introduced for the non-intersective case.

$[\![D]\!]_a{}^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2 . res(W,m) \Downarrow i \in Max_E(t_1) \wedge a_1$ if D is a distributive determiner

$[\![D]\!]_a{}^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2 . res(W,m) \Downarrow i \in Max(t_1) \wedge a_1$ otherwise.

The term a_1 allows for the anaphoric effects of further NPs contained in the nominal, as in the example quoted earlier

(35) No farmers who own a donkey beat it. They are kind to them instead.

The donkeys for each farmer are set up by conditions in a_1.

The remaining step is to make the necessary changes to the syntactic fragment and the translation procedure. Webber's approach was to assign each determiner to one or other class, in effect making the distinction a lexical property. However, this will not do. Some determiners, while tending to be used in one way, can also be used in the other. Example (31), repeated here, shows an intersective use of *few*, and it is also possible (though perhaps harder) to get a non-intersective reading from *most*, as in (125).

(31) Few MPs came to the party, but they (= MPs who came to the party) had a good time.

(125) Most poor farmers only own a donkey. They (= all poor farmers) all want to own a cow as well.

These data suggest that the intersective or non-intersective property should be applied to tokens of determiners, rather than types. The translation procedure must therefore treat determiners as being ambiguous between two cases. It is possible to specify two preference rules which suggest which determiner is more appropriate. The first draws on the presence of other antecedents and anaphors. In an example such as

(126) Few farmers own a donkey. They beat them.

the intersective translation is to be preferred, since there is otherwise no antecedent for *them*. A suitable rule is

Preference rule 3 (non-intersective determiners, I) If a noun phrase is translated as a quantifier Q with index i applied to a set term S, i.e. as a formula of the form $Q(i)\ S$, and S contains a quantifier with index j, and there are anaphors with indices i and j which occur in the same sentence as each other, then prefer the intersective translation for the determiner of the noun phrase.

The second rule makes use of the monotonicity properties of the determiner, defined as follows, from van Eijck (1988):

1. A determiner D is (right) upward monotone if DAB and $B \subseteq B'$ implies DAB'. Examples: *most*, *every*.

2. A determiner D is (right) downward monotone if DAB and $B' \subseteq B$ implies DAB'. Examples: *few*, *no*.

Note that the monotonicity is a truth-conditional property; for example *few* is downward monotone whether functioning intersectively or not for the purposes of anaphora. The second preference rule can then be stated as:

Preference rule 4 (non-intersective determiners, II) If preference rule 3 does not apply, the preferred translation of an upward monotone determiner is the intersective one; of a downward monotone determiner, the non-intersective one.

A justification for the rule can be constructed by observing that downward monotone determiners do not ensure that the intersection is non-empty, while upward monotone determiners do, except in degenerate models. For example, *Few linguists smoke* need not imply that there are any linguists at all who smoke, while *Most linguists smoke* tends to. The non-intersective reading makes an anaphor to a NP with a downward monotone determiner acceptable, while the intersective reading does not guarantee that it will be.

It may be possible to refine this rule further. There is a clear contrast between (124), where the translation is the one predicted by preference rule 4, and (31), where the dispreferred translation is used. The difference between the two sentences is the connective which links the sentences containing the antecedent and the anaphor. *Since* in (124) appears to elaborate on the first sentence, while *but* in (31) creates a contrast. Barwise and Cooper (1981) have suggested than when *but* conjoins noun phrases, it signals that the corresponding quantifiers have opposite monotonicity. A suggestion that follows is to assign a "monotonicity" to sentences, and to specify that words such as *but* signal a change in monotonicity, while *and* and sentence sequencing do not. Preference rule 4 could then be revised as:

Preference rule 5 (non-intersective determiners II, revised) If preference rule 3 does not apply, and the antecedent and anaphor occur in sentences of the same monotonicity, then the preferred translation of an upward monotone determiner is the intersective one; of a downward monotone determiner, the non-intersective one. In sentences of opposite monotonicity the reverse applies.

As it stands, this rule is tentative. For it to be more workable, we would need a better definition of the monotonicity of a sentence, and an indication of which connectives signal a change and why. The applicability of the rule appears to be less clear cut with upward monotone determiners than with downward monotone ones. For example, *most* in general functions intersectively, as the rule would predict. However, in (125), the non-intersective translation is preferred, even though there is nothing to signal it. Similarly, *although* might reasonably be expected to be a connective that signals a change in monotonicity, but in (123), the standard translation still applies. What this may mean is that preference rule 4 should only apply to downward monotone determiners.

There are two final points about non-intersective determiners. Firstly, preference rules 3 and 5 are both triggered by the presence of an anaphor, rather than by the antecedent containing the determiner. This makes them difficult to implement in a fully compositional way. One possible solution is to form both translations of the antecedent when the sentence containing it is processed, assigning different indices to the two quantifiers which result. Preferring the intersective interpretation over the non-intersective one, or vice-versa, then becomes a matter of antecedent resolution, that is of deciding which index to use. The disadvantage in this approach is an increase in processing effort. Secondly, NPs with non-intersective determiners have something in common with generic readings of NPs. This is how they are treated in DRT; Webber also says that she does not make a distinction between non-intersectives and generics, commenting that "the distinction between the set of xs and the generic class of xs is a subtle one" (Webber, 1979, p.2-16). I do not think that the correspondence between the two is exact. On a generic interpretation, the anaphor in

(127) Few farmers own a donkey. They are usually poor.

refers to all prototypical farmers, but not necessarily to all farmers, as it would for a non-intersective reading. A simple adaptation of the interpretation clause given above would place the required information in the DS: instead of taking $Max(t_1)$, simply a "prototypical" sum from t_1 could be used. The process of deciding what makes something prototypical must remain a topic for further investigation.

4.5.2 Disjunction

Disjunction can occur between expressions of many different syntactic categories: sentences (128), noun phrases (129), verb phrases (130), verbs (131) and nouns (132). (Data from Kamp and Reyle.)

(128) The butler loves the baroness or the gardener loves her.

(129) The butler or the gardener loves the baroness.

(130) The butler loathes the gardener or loves the baroness.

(131) The butler loves or admires the baroness.

(132) Every man or woman loves Maria.

Disjunction can also apply to non-constituents (examples adapted from Dowty (1988)):

(133) John will see Mary tomorrow or Bill the day after.

(134) John will give Mary a book or Susan a record.

I will confine attention to some of the forms of constituent disjunction.

Or is sometimes interpreted as inclusive disjunction and sometimes as exclusive disjunction. The examples above can be read either way, with greater or lesser degrees of naturalness. Kamp and Reyle sum up its function as follows:

> The function of **or** is to allow for *alternatives*. The need for alternatives arises in cases where the speaker knows that what he wishes to speak of satisfies one of two or more alternative descriptions, but does not know *which* of those descriptions it satisfies. (p.177)

4.5.2.1 Sentential disjunction

Sentences with anaphora reveal a difference between sentential disjunction and *or* as a truth functional connective. As Kamp and Reyle point out, (135) and (136) read oddly, but (137)–(139) are far more acceptable.

(135) Bill owns a Porsche or Fred owns it.

(136) Jones owns a car or he hides it.

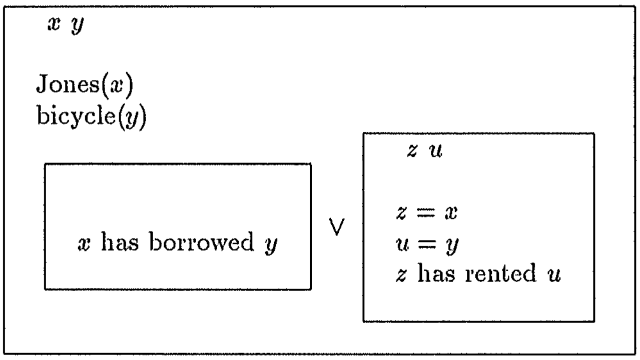
(137) Jones has borrowed a bicycle or he has rented it.

(138) Jones has borrowed a bicycle, or perhaps he has rented it.

(139) Either Jones does not own a car or he hides it.

However, even in the acceptable examples, simple logical disjunction of the sentences is not sufficient to interpret them. Instead, the second sentence has to be interpreted in a context where the antecedents of the first one are accessible, even if the first sentence was false.

Kamp and Reyle's solution in DRT is to construct a DRS for each disjunct. On the basis of (135) and (136), they propose that disjunction renders discourse referents in the disjoined DRSs inaccessible. Kamp and Reyle explain the acceptable (137) and (138) by saying that the two disjuncts provide alternative descriptions of the same event, and it is this which makes the anaphora possible. The events are both about a bicycle which can be identified independently of which event is the actual one. To represent the disjunction in DRT, the discourse referent and condition from *a bicycle* are moved outside the disjoined DRSs, so that the result, for (137), is as follows:

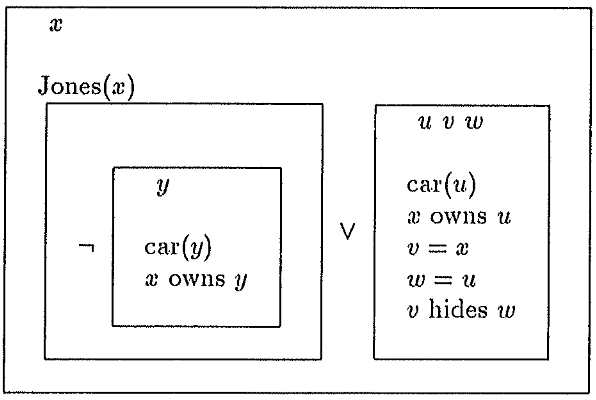


The contrast between (136) and (139) needs a different explanation. Negation normally blocks accessibility in DRT. In (139), accessibility is blocked twice, once by negation and once by disjunction. Kamp and Reyle suggest that such cases often render a discourse referent more accessible than a single blocking, as also happens with double negation. This becomes a little clearer if (139) is paraphrased as (140), which can further be changed to (140a):

(140) Either Jones does not own a car or else/otherwise he hides it.

(140a) Either Jones does not own a car; or it is not the case that Jones doesn't own a car, and he hides it.

Otherwise or *else* make an explicit contrast with the context established by the first disjunct: if it does not hold, the second disjunct must. Kamp and Reyle therefore suggest that the DRS of (139) should take the sub-DRS for the first disjunct, and copy its negation into the sub-DRS for the second one. In doing so, the negation present because of *doesn't own...* is cancelled out, making the discourse referent for *a car* accessible. The final DRS looks like this:



In (136), the same mechanism can apply, but this time there is no negation in the first sentence to cancel out, and so *a car* is inaccessible in the copied DRS.

Kamp and Reyle's account appears to work well. The major weakness in it is that it does not provide explicit conditions for when copying is to be carried out, and when moving the condition outwards, as in (137), is appropriate. It may well be that no definitive answer can be given. Certainly there are some syntactic constructs, such as *either...or...*, which encourage one mechanism rather than another, but there are other factors too. For example, (135) can be made acceptable using the same mechanism as (137), when some particular car is identified. Kamp and Reyle suggest this might happen with a NP that is very specific, such as "a Porsche which I have seen race past our house several times this morning." Other sentences in the discourse may also contribute:

(141) Bill owns a Porsche or Fred owns it. Whichever one it is, I just saw him run over a toad in it.

Something similar was observed for donkey sentences, particularly (48) of section 2.3.1.

What (137) and (141) depend on is the interpretation of the NPs as referring to something specific. The acceptability of (128), where the NPs are definite, is in keeping with this idea. So another solution might be to employ a mechanism based on something resembling presupposition. In effect, we say that to interpret (137), presuppose that there is something specific for *a bicycle* to refer to, and then treat the occurrence of *a bicycle* in the first disjunct as being anaphoric to this thing. A mechanism for presupposition in DRT has been proposed by van der Sandt (1990), in which the conditions that have to be accommodated to interpret a sentence are moved outwards from the DRS where they arise to a surrounding DRS, in a manner very similar to Kamp and Reyle's solution to (137). The same approach can be applied to (135)–(138), provided there is something which allows the presupposition to be made. (139) is a little harder to deal with in the same way, but it is possible. A paraphrase of how to interpret the sentence might be:

(142) For any car: presuppose that is it relevant to the discourse. Then either Jones doesn't own it, or he hides it.

A similar approach is taken in TAI. Simply forming the disjunction of the interpretations of the formulae involved will not work, since if the first formula is false, then the DS will not contain information which the anaphor in the second can pick up. What we want to do is find a way of getting the content information from the NPs of the first disjunct into the second disjunct. If either disjunct is unsatisfied, then no dependence information must be set up from it. One way of doing this is to use the widening rule of section 4.4.5, in such a way that the anaphor appears within the scope of the widened quantifier. A suitable rule is:

Existence presupposition rule.

Given a formula containing a quantifier Q with index i derived from an antecedent, and a quantifier Q' with input index i derived from an anaphor, apply the widening rule to Q in such a way that the result is $Q\ \hat{u}[S]$, where Q' occurs within S.

The quantifier which is widened may be of any kind, including (translations of) proper nouns and anaphors as in (137)–(139). The same mechanism will allow the conditional example (97) of section 4.3.5 (repeated here)

(97) If Bill/a (specific) farmer owns a donkey, he beats it. He is cruel.

to be dealt with, by widening the subject of the antecedent sentence to outside the conditional.

Formation rule F9-or adds a disjunction operator to L(GQA).

F9-or. If F_1 and F_2 are formulae, then $F_1 \bigvee F_2$ is a formula.

Interpretation:

$$[\![F_1 \bigvee F_2]\!]_t{}^{W,m} = [\![F_1]\!]_t{}^{W,m} \vee [\![F_2]\!]_t{}^{W,m}$$
$$[\![F_1 \bigvee F_2]\!]_a{}^{W,m} = [\![F_1]\!]_a{}^{W,m} \vee [\![F_2]\!]_a{}^{W,m}$$

The meta-language operator $\vee$ is the standard Boolean inclusive-or. An exclusive version could trivially be formulated as well. The changes to the English fragment and the translation procedure are:

Syntactic categories

Conj (conjunction)

Production rules

S16. D:cf_1f_2 $\rightarrow$ S:f_1 Conj:c S:f_2

Extended versions of S16 for three or more sentences could also be added, without introducing any new principles of interest. Using $Pre(f)$ as shorthand for the result of applying the existence presupposition rule zero or more times to f, the lexical translation c of *or* is $\lambda f_1.\lambda f_2.Pre(f_1 \bigvee f_2)$. As an example, for (137), the translation before applying *Pre* is

(143) $((the\ jones)(i)\ \hat{x}[(\Phi\ 1(bicycle_s))(j)\ \hat{y}[borrow(x,y)]]) \bigvee$
$((the\ Of(i,sg))(i)\ \hat{x}[(the Of(j,sg))(j)\ \hat{y}[rent(x,y)]])$

eventually leading to

(144) $(the\ jones)(i)\ \hat{u}[(\Phi\ 1(bicycle_s))(j)\ \hat{v}[$
$((the\ Of(i,un))(i)\ \hat{x}[(the\ Of(j,un))(j)\ \hat{y}[borrow(x,y)]]) \bigvee$
$((the\ Of(i,sg))(i)\ \hat{x}[(the Of(j,sg))(j)\ \hat{y}[rent(x,y)]])]]$

The formula puts Jones into slot i and a bicycle into slot j, and then interprets the disjuncts to be consistent with these slots, so that slot i borrows slot j or slot i rents slot j.

As it stands, the existence presupposition rule is applied blindly, and it should be more tightly constrained: it need not apply to all quantifiers in a sentence. For example, in

(145) Jones has borrowed a bicycle from a friend or he has rented it from a shop.

neither a friend nor a shop should not be presupposed. Presupposing all NPs to which there is an anaphor is not right either. For example, in (146),

(146) Every farmer who owns a donkey beats it or John walks.

the disjunction would trigger application of the presupposition rule to *a donkey*, since there is an anaphor to it. However, there is no need to presuppose a particular donkey, and doing so in fact leads to an unlikely reading of the sentence. A possible rule for *Pre* is therefore: given a formula $F_1 \bigvee F_2$, if there is a quantifier $Q(i)$ in F_1 and an anaphoric quantifier $(\mathit{the}\ \mathit{Of}(i, P))(j)$ in F_2, then presuppose $Q(i)$. The rule can be applied zero or more times. The major thing that is missing from this account is an explanation of why the presuppositions are easier in (137) and (138) than (135) and (136). As I commented above, there appear to be various complex factors of context behind this. The answer (and indeed the data) is not clear, and I think is likely to require some appeal to pragmatic issues, of what is being communicated, and what is needed to make the communication effective.

For *either...or...* disjunctions of sentences, we do not want to make an existence presupposition which holds outside the disjoined sentence, but to say that if such a presupposition is made, then the disjunction holds. The way to do this is at the level of the discourse grammar. The solution is similar to that adopted for symmetric readings of conditionals and sentences related by adverbs of quantification: when interpreting with respect to a set of DSs V, pass on either V unchanged or the empty set, depending on whether the translation of the *either...or...* sentence can be satisfied. Formation rule D4 is added to L(GQAD):

D4. If F is a formula, then $\mathit{Hyp}(F)$ is a formula.

Interpretation: $I(\mathit{Hyp}(F), V) = \{W \in V \mid I(F, V)\}$

$\mathit{Hyp}(F)$ can be read as "hypothesise F". In the syntax, *either...or...* is introduced syncategorematically.

S17. D:$\mathit{Hyp}(\mathit{Pre}(f_1 \bigvee f_2)) \rightarrow$ either S:f_1 or S:f_2.

One minor point concerns the parallel between widening as used for quantifier scoping and as used for existence presupposition. In

(147) Every farmer owns a donkey or some merchant owns it.

the second disjunct can be given a reading in two ways. Either the disjunction can trigger the existence presupposition rule as above, or the quantifier scoping rule can widen *a donkey* over *every farmer* for some independent reason such as background knowledge. The resulting TAI expressions are the same in the two cases.

There is one more standard "difficult" example:

(148) Either there is no bathroom in this house, or it is in a funny place.

I will return to (148) following the discussion of negation in section 4.5.4.

4.5.2.2 Non-sentential disjunction

Kamp and Reyle's treatment of non-sentential disjunction is to "multiply out", so that (149) is analysed as if it were (149a).

(149) Mary loves Bill or hates Fred.

(149a) Mary loves Bill or Mary hates Fred.

Multiplying out will work for most syntactic categories, although it must be carried out after initial conversion to a logical representation. For example,

(150) A woman loves Bill or hates Fred.

does not mean the same as:

(151) A woman loves Bill or a woman hates Fred.

but if a discourse referent is formed for *a woman*, and then used in the multiplied out sentence, the result is correct. Two exceptions to multiplying out noted by Kamp and Reyle are nouns and noun phrases. They point out that

(152) Every man, woman or child admires Maria.

is not equivalent to

(153) Every man admires Maria, or every woman admires Maria or every child admires Maria.

but that the following paraphrase does work:

(154) Everyone who is a man, who is a woman or who is a child admires Maria.

The DRS constructed from the sentence contains a discourse referent x, and a disjunction of sub-DRSs, each applying one of the noun conditions to x.

A more direct approach is used in TAI. L(GQA) is first extended with a set union operator, defined for both basic predicates and for set terms. The disjunction symbol is used, "overloading" it, in the sense that it may apply to expressions of more than one type.

F10-or. If R_1 and R_2 are n-place predicate symbols, then $R_1 \bigvee R_2$ is an n-place predicate.

Interpretation:

$$\begin{aligned}
[\![R_1 \bigvee R_2]\!]_t{}^{W,m} &= [\![R_1]\!]_t{}^{W,m} \cup [\![R_2]\!]_t{}^{W,m} \\
[\![R_1 \bigvee R_2]\!]_a{}^{W,m} &= [\![R_1]\!]_a{}^{W,m} \vee [\![R_2]\!]_a{}^{W,m}
\end{aligned}$$

F11-or. If S_1 and S_2 are set terms, then $S_1 \bigvee S_2$ is a set term.

Interpretation:

$$\begin{aligned}
[\![S_1 \bigvee S_2]\!]_t{}^{W,m} &= \lambda i.([\![S_1]\!]_t{}^{W,m} i) \cup ([\![S_2]\!]_t{}^{W,m} i) \\
[\![S_1 \bigvee S_2]\!]_a{}^{W,m} &= \lambda i.([\![S_1]\!]_a{}^{W,m} i) \vee ([\![S_2]\!]_a{}^{W,m} i)
\end{aligned}$$

Disjunctions of verb phrases, verbs and nouns are then easily formulated.

S18. VP:cp_1p_2 $\rightarrow$ VP:p_1 Conj:c VP:p_2

S19. TV:cp_1p_2 $\rightarrow$ TV:p_1 Conj:c TV:p_2

S20. DV:cp_1p_2 $\rightarrow$ DV:p_1 Conj:c DV:p_2

S21. N$'$:cp_1p_2 $\rightarrow$ N$'$:p_1 Conj:c N$'$:p_2

For *or*, c is as specified above, i.e. $\lambda f_1.\lambda f_2.Pre(f_1 \bigvee f_2)$. An example showing why the existence presupposition rule is needed in S18–S21 is the VP disjunction in

(155) Mary has borrowed a bicycle or has rented it.

A bicycle has to be presupposed to give a reading to the second disjunct.

The difficulty that arises with NPs is illustrated by (156) and (157).[8]

(156) Bill or John is reading Ulysses. He doesn't like Anna Karenina.

(157) The lead singer of Thröttler plays a chainsaw or a power drill. It makes an ungodly racket.

In (156), the anaphor refers to either Bill or John, regardless of which is reading Ulysses, and in (157) to the instrument, whatever it turns out to be. The initial version of the DRS construction algorithm gives a DRS which can be paraphrased as: there is an object x such that either $Bill(x)$ holds or $John(x)$ holds, and also "x is reading Ulysses" holds, and the pronoun is translated as x. However, this forces the subject of the second sentence to be whoever is reading Ulysses – exactly what we wanted to avoid. The DRT solution is to introduce three discourse referents, one for the object that reads Ulysses x, one for Bill y and one for John z. One of the disjoined sub-DRSs equates x with y and the other x with z. A similar mechanism is used for (157).

For NP disjunction, the $\bigvee$ operator is extended to quantifiers. Following Barwise and Cooper (1981, p.194), the interpretation of a disjoined generalized quantifier is the union of the disjuncts' interpretations. Each of the quantifiers has its own index, as does the quantifier formed from them. The anaphoric information passed on by the combined quantifier is the sum of the information from each quantifier separately, and its structure is formed by merging the structuring information of the disjuncts. Thus in

(158) Some men or some women each went to a party.

on the reading where *or* as taken as inclusive, and each individual went to a different party, we can say

(159) They enjoyed them.

to mean that each individual enjoyed the party he or she went to.

The new rule for L(GQA) is:

F12-or. If Q_1 and Q_2 are quantifiers and i_1 and i_2 are indices, then $Q_1(i_1) \bigvee Q_2(i_2)$ is a quantifier.

$$\begin{aligned}
[\![Q_1(i_1) \bigvee Q_2(i_2)]\!]_t{}^{W,m} &= \lambda i.([\![Q_1]\!]_t{}^{W,m} i_1) \cup ([\![Q_2]\!]_t{}^{W,m} i_2) \\
[\![Q_1(i_1) \bigvee Q_2(i_2)]\!]_a{}^{W,m} &= \lambda i \lambda t \lambda a.(([\![Q_1]\!]_a{}^{W,m} i_1\ t\ true) \vee ([\![Q_2]\!]_a{}^{W,m} i_2\ t\ true)) \\
&\qquad \wedge a \wedge Merge(W, i, i_1, i_2)
\end{aligned}$$

true is passed to the quantifiers rather than a, since a does not depend on the quantifiers or their indices. The operation *Merge* is defined as

[8] Kamp and Reyle's equivalent to (156) uses Mary rather than John, in which case gender agreement makes it clear which the antecedent is.

$$
\begin{aligned}
Merge(W,i,j,k) \;=\; & (W\Downarrow i = W\Downarrow j \oplus W\Downarrow k) \wedge \\
& \forall X \in W\downarrow j[W\backslash(X,i)=\emptyset \vee W\backslash(X,i)=W\backslash(X,j)] \wedge \\
& \forall X \in W\downarrow k[W\backslash(X,i)=\emptyset \vee W\backslash(X,i)=W\backslash(X,k)]
\end{aligned}
$$

That is, the i slot is equal in content to the sum of the j and k slots. For everything that appears in the j slot, if it also appears in the i slot, it stands in the same dependence relations in the two, and similarly for the k slot. If one of the disjuncts – k, say – was unsatisfied, then $W\downarrow k$ is empty, and so it does not constrain slot i.

The syntax rule is:

S22. NP:$c(q_1(i_1))(q_2(i_2)) \rightarrow$ NP:q_1 Conj:c NP:q_2

The indices i_1 and i_2 are assigned uniquely, in the usual way. Returning to two earlier examples, in (156) reference can be made to either *Bill* or *John* regardless of who the reader is by setting the anaphor's input index to that of the appropriate disjunct. It is necessary to move the quantifiers out of the disjunction by application of the existence presupposition rule as well, just as it is necessary to move proper names to top level in DRT. For (157), the anaphor's input index is that of the disjoined quantifier as a whole – it does not matter which of the disjuncts satisfied the truth condition.

4.5.3 Conjunction

Conjunction has similar syntactic distribution to disjunction, but is a little simpler semantically. Sentential conjunction can be handled using the sequencing rule of section 4.3.5, or alternatively by introducing a conjunction operator on L(GQA) formulae:

F9-and. If F_1 and F_2 are formulae, so is $F_1 \bigwedge F_2$.

Interpretation:

$$
\begin{aligned}
[\![F_1 \bigwedge F_2]\!]_t{}^{W,m} &= [\![F_1]\!]_t{}^{W,m} \wedge [\![F_2]\!]_t{}^{W,m} \\
[\![F_1 \bigwedge F_2]\!]_a{}^{W,m} &= [\![F_1]\!]_a{}^{W,m} \wedge [\![F_2]\!]_a{}^{W,m}
\end{aligned}
$$

Syntax rule S16 can be used, with the lexical translation of *and* being $\lambda f_1.\lambda f_2.f_1 \bigwedge f_2$.

For conjunction of verbs, verb phrases, nouns and noun phrases, the rules of the previous section may be used, with the substitution of $\bigwedge$ for $\bigvee$, $\cap$ for $\cup$ and $\wedge$ for $\vee$. No change is needed to the syntax rules. Hence:

F10-and. If R_1 and R_2 are n-place predicate symbols, then $R_1 \bigwedge R_2$ is an n-place predicate.

Interpretation:

$$
\begin{aligned}
[\![R_1 \bigwedge R_2]\!]_t{}^{W,m} &= [\![R_1]\!]_t{}^{W,m} \cap [\![R_2]\!]_t{}^{W,m} \\
[\![R_1 \bigwedge R_2]\!]_a{}^{W,m} &= [\![R_1]\!]_a{}^{W,m} \wedge [\![R_2]\!]_a{}^{W,m}
\end{aligned}
$$

The semantics already contains a rule for $S_1 \bigwedge S_2$, which was used for relative clauses.

For quantifiers, the formation rule is F12-and.

F12-and. If Q_1 and Q_2 are quantifiers and i_1 and i_2 are indices, then $Q_1(i_1) \bigwedge Q_2(i_2)$ is a quantifier.

The interpretation is a little more subtle than the other cases. The conjunction can be either distributive or collective. Example (160) shows the difference.

(160) Two men and two women met.

The sentence can mean either that there was a meeting of two men and a meeting of two women, or one or all four individuals, or various intermediate cases, such as two meetings each involving one man and one woman. A suitable truth conditional clause is

$$[\![Q_1(i_1) \bigwedge Q_2(i_2)]\!]_t^{W,m} = \lambda i.\{X \subseteq \mathcal{D} \mid \exists Y \in ([\![Q_1]\!]_t^{W,m} i_1) \cap ([\![Q_2]\!]_t^{W,m} i_2)[Sup(^*Y) \in X]\}$$

That is, we find a set Y which contains things satisfying both quantifiers (and possibly some other entities), and ensure that everything in it, after sum closure is also in X. The translation and interpretation of (160) are (after simplifying):

(160a) $((2_= \ man_s)(j) \bigwedge (2_= \ woman_s)(k))(i) \ met_s$

(160.T) $\exists Y \subseteq \mathcal{D}[2'_=(man^*, Y) \wedge 2'_=(woman^*, Y) \wedge Sup(^*Y) \in met^*]$

Supposing the only meeting is between all four individuals, so that met^* contains $m_1 \oplus m_2 \oplus w_1 \oplus w_2$ but no sub-sums of it. Then a Y which satisfies both quantifiers is $\{m_1 \oplus m_2, w_1 \oplus w_2\}$, the supremum of the sum closure of which is precisely this member of met^*. The anaphoric part is a little harder. If it were not for collective readings of the conjoined quantifier, we could use something similar to disjunction. Instead, we must set slot i_1 to be all the things which satisfy the set term of Q_1 and which are part or all of something satisfying t (the VP extension in (160)); similarly for i_2. Slot i is the result of merging the slots i_1 and i_2. Finally, the whole of slot i must satisfy t. Hence:

$$\begin{aligned}[\![Q_1(i_1) \bigwedge Q_2(i_2)]\!]_a^{W,m} &= \lambda i \lambda t \lambda a.([\![Q_1]\!]_a^{W,m} i_1 \ t' \ true) \wedge ([\![Q_2]\!]_a^{W,m} i_2 \ t' \ true) \wedge \\ &\qquad W \Downarrow i \in t \wedge a \wedge Merge(W, i, i_1, i_2) \\ \text{where } t' &= \{X \in \mathcal{D} \mid \exists Y \in t[Ats(X) \subseteq Ats(Y)]\}\end{aligned}$$

t' is the collection of all sub-sums of things in t. The occurrences of *true* just reflect the fact that we get the anaphoric restriction from the VP through a, as for disjunction. For (160), we have:

(160.A) $W \Downarrow j \in Max(man^* \cap T) \wedge W \Downarrow k \in Max(woman^* \cap T) \wedge W \Downarrow i \subseteq met^* \wedge Merge(W, i, j, k)$

where $T = \{X \in \mathcal{D} \mid \exists Y \in met^*[Ats(X) \subseteq Ats(Y)]\}$

meaning that even if it is the case that all four individuals met, $W \Downarrow j$ is just the men who met, and so on. If instead of *met*, we had *own a donkey*, then it would be the i slot that stood in relation to the donkeys owned, and if there are some men who only owned donkeys jointly with a woman, the definition allows there to be no dependence between slot j and the "donkey" slot for such individuals.

NP conjunction in TAI contrasts with the approach in the Kamp and Reyle version of DRT, where there is an ambiguity between the collective and distributive readings. The collective reading is formed first, an optional rule then forming the distributive reading from it.

One cause for concern is whether *and* should be translated as conjunction for all the lexical types it occurs with. Example (161),

(161) Most men walk and talk.

seems to mean that most men both walk and talk, which is consistent with intersecting the set terms from the verbs. However, with noun conjunction, the interpretation is usually more like disjunction:

(162) Most men and women talk.

(162) means that most things which are either men or women talk. There are noun conjunctions where the intersection is more appropriate, as in

(163) Ronald Reagan was a president and actor.

but they appear to be rarer than the disjunctive interpretation. A variant of rule S21 may be needed for this case:

S21.N. N′:$p_1 \bigvee p_2$ $\rightarrow$ N′:p_1 and N′:p_2

L(GQA) makes it possible to deal with a problematic example from Link (1987):

(164) John and Mary invited their[distr] parents to their[coll] house.

Link's reading of the sentence is that John invited John's parents to John and Mary's house, and Mary invited Mary's parents to John and Mary's house. His solution is to translate the VP in terms of a free variable z, to which the sum $john \oplus mary$ must be bound, and a lambda-abstracted variable u, which is distributed over the members of the same sum. Link gives no hints about how this is to be achieved compositionally. As it stands, L(GQA) does not include a mechanism for possessives. Assuming a predicate $parentof(a, b)$, which holds if a are the parents of b, then *their parents* can be translated as:

$$(the\ \hat{u}[(the\ Of(i, pl))(i)\ \hat{v}[parentof(u, v)]])$$

which means the us which are parents of slot i of the DS. Employing a similar predicate for *their house*, the sentence translates as

(164a) $((the\ john)(f) \bigwedge (the\ mary)(g))(i)$
$\hat{x}[(the\ \hat{u}[(the\ Of(i, pl))(i)\ \hat{v}[parentof(u, v)]])(j)$
$\hat{y}[(the\ \hat{p}[(the\ Of(i, pl))(i)\ \hat{q}[houseof(p, q)]])(k)$
$\hat{z}[invite(x, y, z)]]]$

The formula says, roughly, that John and Mary appear in slot i, and that the parents of slot i are a y such that the house of slot i is a z such that John and Mary invite y to z. In the interpretation, the extensions of *parentof* and *houseof* will be closed under summation, so, for example $parentof(u, v)$ is true if u is Mary's parents and v is Mary, and also if v is Mary and John together and u is their parents together. The formula is hence true on either the collective or the distributive reading of either of the *their*'s, including Link's interpretation.

4.5.4 Negation

Like disjunction and conjunction, negation can apply to a number of different syntactic constructions. Some forms of negation may give rise to more than one interpretation. For example,

(165) A man does not own a donkey.

could mean that there is no man who owns a donkey, or that the speaker has a man in mind but he owns no donkeys. By placing emphasis on *own*, a reading where a man and a donkey stand in some relation other than ownership can also be obtained. I will call the three forms of negation sentence negation, VP negation and verb negation. Sentence negation can be obtained, perhaps more in the speech of logicians than average language users, by prefixing a sentence with *it is not true that* or *it is not the case that.* Negation of determiners is also possible:

(166) Not many men own a donkey.

Negation of nouns is harder, although prefixing the noun with *non* has some sort of negating effect. I will confine my attention to sentence and verb phrase negation.

Anaphora to a NP within the scope of a negation is not generally possible, as the following examples show:

(167) A man does not own a donkey. ?It brays. ?He owns a cow.

(168) It is not true that a man owns a donkey. ?It brays. ?He owns a cow.

DRT takes a structural approach to negation, making discourse referents within a negated DRS inaccessible. DPL also blocks anaphora, by cancelling bindings to variables made within a negation. However, there are a number of special cases. Firstly, double negation generally seems to restore antecedents, as in

(169) It is not true that John doesn't own a car. It is bright red and parked outside. (Groenendijk and Stokhof, 1991b, p.91)

Even where there is single negation, anaphora is sometime possible, as in Kamp and Reyle's

(170) Jones doesn't like a Porsche. He owns it.

which can be read as saying there is a specific Porsche which Jones does not like, but which he owns. Something similar can make the last sentence of (168) acceptable, if a specific man or donkey is meant. Kamp and Reyle do not give a solution to this problem. For TAI, the existence presupposition rule can be allowed to apply to *a Porsche*, so that it is widened to outside the negation. They also quote

(171) Jones doesn't own Ulysses. He likes it (however).

as an acceptable example, on the grounds that DRT moves proper nouns to the top-level DRS from which they are always accessible. I am unconvinced by this argument – to me, the example can be paraphrased as "Jones doesn't own a copy of Ulysses. He likes the (abstract entity which is the) novel Ulysses (however)." That is, the actual antecedent of the anaphor is obtained by a bridging assumption from the NP *Ulysses*. The mechanism behind this may be related to

(172) The ham sandwich at table three wants a coke. He's getting impatient.

quoted in section 2.5.4.

Another special case, discussed by van den Berg (1990), is

(173) It is not true that there is a man walking in the park. They are all home watching TV.

He uses this example to motivate a variant of DPL in which a noun phrase passes on a verifier and a falsifier, i.e. the assignments which make the sentence within the negation true and false respectively. Anaphors are interpreted with respect to the verifier. Negation has the effect of interchanging verifier and falsifier. This approach, or something equivalent to it, does appear to be correct, and should generalise to determiners other than the ones he considers (*a*, *every*, *some* and *no*). For example, in

(174) It isn't true that most men are home watching TV. They are walking in the park.

the second sentence has a reading that all men who are not watching TV are in the park. Adding *instead* helps obtain this reading. There is possibly an argument for a reading of the second sentence as meaning all men are walking in the park. However, I think this need not be made distinct from the other: it is just an extreme case, where there are no men watching TV.

TAI is easily adapted to sentence negation. The semantic account of acceptability will deal with (167)–(169), since the anaphoric information provided by the antecedent is empty. The changes to L(GQA) and the syntactic fragment are, where $\sim$ is Boolean negation in the meta-language:

F13. If F is a formula, then $\neg F$ is a formula.

Interpretation

$$[\![\neg F]\!]_t{}^{W,m} = \sim [\![F]\!]_t{}^{W,m}$$
$$[\![\neg F]\!]_a{}^{W,m} = [\![F]\!]_a{}^{W,m}$$

To understand how this works, consider the sentence negation of

(175) Some men do not own a donkey.

presented as (176).

(176) $\neg((some\ man)(i)\ \hat{x}[(\Phi\ 1(donkey_s))(j)\ \hat{y}[own(x,y)]])$

(176.T) $\sim (some'(man^*, oad))$

(176.A) $W \Downarrow i \in Max_E(man^* \cap oad) \wedge$
$W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W \backslash (X,i) \Downarrow j \in Max(donkey^* \cap own^*(X)) \wedge$
$W \backslash (X,i) \Downarrow j \subseteq own_\perp{}'(X)\}$

where $oad = {}^*\{X \in \mathcal{D} \mid \Phi'(1' \cap donkey^*, own^*(X))\}$
The truth condition says that it is not the case that there are any men who own a donkey. The first line of the anaphoric part sets slot i to all men who own a donkey. There are no such men, and so the content of the slot is $\perp$. By a similar reasoning to that employed for the example with determiner *no* in section 4.3.3, slot j must also be empty.

There are two possible interpretations for verb phrase negation. One reading is that there is a known group of men, who do not own any donkeys, and a subsequent anaphor will refer to this group of men. On the other reading, an anaphor picks up all men who do not own a donkey. The first case is best analysed as sentence negation, with "some men" widened to outside the negation using the existence presupposition rule. For the second case, the changes to L(GQA) are straightforward. The operator $-$ means complement with respect to $\mathcal{D}$.

F14. If S is a set term, then $\neg S$ is a set term.

Interpretation

$$\begin{aligned} [\![\neg S]\!]_t^{W,m} &= -[\![S]\!]_t^{W,m} \\ [\![\neg S]\!]_a^{W,m} &= [\![S]\!]_a^{W,m} \end{aligned}$$

The verb phrase negation of (175) will then be:

(177) $(some\ man)(i)\ \neg(\hat{x}[(\Phi\ 1(donkey_s))(j)\ \hat{y}[own(x,y)]])$

(177.T) $some'(man^*, -oad)$

(177.A) $W{\Downarrow}i \in Max_E(man^* \cap -oad) \wedge$
$W{\downarrow}i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X,i){\Downarrow}j \in Max(donkey^* \cap own^*(X)) \wedge$
$W\backslash(X,i){\Downarrow}j \subseteq own_\perp'\}$

where $oad\ =\ {}^*\{X \in \mathcal{D} \mid \Phi'(1' \cap donkey^*, own^*(X))\}$
The truth condition is that there are some men, each of whom does not own any donkeys. The anaphoric condition set slot i to these men. For each man in slot i, the contents of slot j is $\perp$.

The additional syntax rules are:

S23. S:$Pre(\neg(f))$ $\rightarrow$ IINTT S:f

S24. S:$Pre(\neg(qs))$ $\rightarrow$ NP:q do not VP:s

S25. VP:$Pre(\neg(s))$ $\rightarrow$ do not VP:s

Rule S24 and S25 allow for the ambiguity between sentence and VP negation. *Do* must be changed to the appropriate morphological variant. *IINTT* is shorthand for the negation phrases mentioned above, such as *it is not the case that*. *Pre*, which allows the existence presupposition rule to apply, is used to give a reading to sentences such as (170).

4.5.4.1 Negation and disjunction

The example

(148) Either there is no bathroom in this house, or it is in a funny place.

was mentioned above as a difficult case of negation. It is possible to analyse this sentence if the determiner *no* is treated as being equivalent to, roughly, *not a*, just as in first order logic, it is equivalent to $\neg\exists x$. For TAI, if *a* in *not a* is treated as $\lambda s.(\Phi\ 1(s))$, then $(noP)(i)\ Q$ becomes $\neg((\Phi\ 1(P))(i)\ Q)$. For (148) the translation would be:

$$Hyp((\neg((\Phi\ 1(bath_s))(i)\ inhouse_s))\ \bigvee\ ((the\ Of(i,sg))(i)\ infunny_s))$$

where *Hyp* is the L(GQAD) operation defined in section 4.5.2. After applying the existence presupposition rule, the result is

$$Hyp((\Phi\ 1(bath_s))(i)$$
$$\hat{u}[(\neg((the\ Of(i,un))(i)\ inhouse_s))\ \bigvee\ ((the\ Of(i,sg))(i)\ infunny_s))$$

Ignoring the effect of *Hyp*, the interpretation of this formula is:

(148.T) $\Phi'(1' \cap bath^*, {}^*\{U \in \mathcal{D} \mid \sim (W \Downarrow i \in inhouse^*) \ \vee\ W \Downarrow i \in infunny^*\})$

(148.A) $W \Downarrow i \in Max(1' \cap bath^*) \cap^* \{U \in \mathcal{D} \mid \sim (W \Downarrow i \in inhouse^*) \ \vee\ W \Downarrow i \in infunny^*\}) \wedge$
$W \downarrow i \subseteq \{U \in \mathcal{D}_\perp \mid W\backslash(X,i) \Downarrow i \in Max(\sharp(W\downarrow i) \cap inhouse^*) \ \vee$
$W\backslash(X,i) \Downarrow i \in Max(\sharp(W\downarrow i) \cap infunny^*)\} \wedge$
$sg'(W \Downarrow i)$

In the truth conditional part and the first line of the anaphoric part, $\{U \ldots\}$ is either empty or is the whole domain $\mathcal{D}$. So the truth conditions amount to there being some object in $\mathcal{D}$ which is a bathroom, and $W \Downarrow i$ meeting the conditions specified in the body of the set term. Suppose that there is no bathroom in the house. Then $\sim (W \Downarrow i \in inhouse^*)$ is true for any bathroom in $W \Downarrow i$, and the truth conditions can be met. Alternatively, suppose there is a bathroom in the house, but it is in a funny place. Then $W \Downarrow i \in infunny^*$ holds if $W \Downarrow i$ is this bathroom, and again the truth conditional part can be met. Similar reasoning shows that the anaphoric conditions will be met for this $W \Downarrow i$. Finally, the *Hyp* operator means that the DS will be reset after processing the sentence, in the sense that slot i will be left as it was before.

4.6 Summary

The chapter started with a conceptual model for noun phrase anaphora, specifying all the information from the antecedent which must be available to interpret the anaphor. An extensional representation of the information was designed, and a logic which extracts the representation from a formula devised. The conceptual model and representation each capture part of the empirical content of the theory. The logic, in contrast, is essentially a formal device. The remaining empirical content is contained in the translation procedure from natural language sentences to formulae of the logic. The logic is completely deterministic, while the translation procedure allows more than one translation to be derived from the same parsed sentence. The flexibility that this makes available is required to capture the variable intuitions about the data described in chapter 2. A potential disadvantage with this degree of flexibility is that of weakening the empirical content of the theory. Preference principles go some way towards rectifying this. Finally, a sketch of three logical connectives was added to the theory, again introducing some ambiguity into the range of available translations.

The meta-language expressions obtained from interpreting the logic appear to be complex, which has the potential disadvantages of obscuring insights into the linguistic properties of anaphora, and of being computationally intractable. I have no answer to the first of these points. Ideally, a simpler logic should be found, if it is possible to do so; my criterion has been more one of truth than of beauty. The second point is addressed in section 5.4, where it is shown that efficient computation with the meta-language expressions is possible. One other property of the logic is that, despite its complexity, it is relatively conventional, being static and based in standard set theory and, for the plural representation, lattice theory. This compares favourably with the theories of anaphora of chapter 3. DRT has a conventional logic, but relies on representational devices which are derived in an unconventional way, and which are an essential part of the empirical content of the theory. Dynamic logics do not rely on the representation, but do require non-standard logics, the properties of which are still a matter of investigation, and for which theorem proving techniques have yet to be developed.

5 Evaluation

In this chapter, I evaluate various aspects of TAI. Firstly, the relation with DRT and dynamic logics is explored. Section 5.2 examines two methodological aspects of TAI: how to construct a theory which is flexible enough to capture all the required readings, without falling into the trap of being too vague, and whether the representation of context is appropriate. The following section is concerned with the empirical coverage of TAI, and sketches some extensions to the theory. Two computational aspects of TAI are dealt with in section 5.4: how it relates to Webber's theory of definite anaphora, and how DSs may be computed efficiently. The final section summarises the major points of the discussion and suggests how research might continue.

5.1 Comparison with other theories

5.1.1 The conceptual model and its formalisation

In comparing TAI with the dynamic logics of anaphora and with DRT, there are several points of contact: how they each relate to the conceptual model of context, how the context is formally represented, and how accessibility of antecedents to anaphors is defined. In this section, the terms of section 4.2 are used to compare the theories.

TAI

To recap the conceptual model behind TAI: the anaphoric information is contained in a context, consisting of a number of slots labelled with indices. Each slot has both content (the referent of the NP) and structure (how it relates to other antecedents). The latter takes the form of dependence relations between the content of two or more slots, expressed as a partitioning of the slot. A NP which can act as an antecedent for both singular and plural anaphors, as in

(1) Most^i farmers who own a^j donkey beat it^j. They^j are looking for new masters.

is represented uniformly for the two cases. That is, there is a single context, containing all the farmers in one slot and all the donkeys in another. *They* is interpreted as the whole of the donkey slot. *It* is interpreted with respect to the same context, but using the identity of each farmer in turn to isolate just the relevant part of it.

The particular formalisation of the conceptual model used in L(GQA) is *total* and *extensional*. The former means that the context, in the form of a DS, always has a slot for every NP, even before the NP has appeared in the discourse, and the entire context can be seen as being defined by a conjunction of constraints upon it.[1] DSs are extensional in the sense that all of the information about the slots is explicitly enumerated, resulting in the need for devices such as the empty entity ($\perp$). Indeterminacy is represented formally, in working with sets of possible DSs at the discourse level. The same mechanism, of multiple contexts, is used for conditionals.

Finally, TAI represents the accessibility of antecedents largely in semantic terms, with one additional principle, the PVTP, which is stated in structural terms. In general, an antecedent is inaccessible to an anaphor if the information in the antecedent's slot of the context does not meet the semantic number requirements of the anaphor.

[1] The term *total* is intended to contrast with the case where the anaphoric information is divided between different parts of the semantic representation, rather than as in "total" and "partial" functions.

Dynamic logics

In DPL, the context is represented as an assignment function, with variables acting as the labels on slots. Because DPL only deals with singular anaphora, the issues of partitioning and dependence do not arise. For both quantified and conditional sentences, something equivalent to multiple contexts is used. For example, the evaluation of a donkey sentence (of either sort) involves constructing the context for each farmer and donkey in turn. The treatment of indeterminacy, cases where there is not enough information to identify a single referent, is informal, although employing sets of possible assignments would overcome this.

DPLP, the extension of DPL to plurals by van den Berg (1990), represents plural referents by means of multiple contexts. An expression is interpreted with respect to a set of assignments, each assignment mapping a variable to one of the individuals which together form the referent of the plural noun phrase. Dependence relations then take the form of two individuals occurring in the same context. There is no way of violating the dependence relation to obtain the less prominent reading in sentences like

(2) Every^i farmer owns a^j donkey. They^i beat them^j.

The slots still contain individuals, but there is no reason in principle why they could not contain collections (sums) of individuals for examples such as (3):

(3) Three^i boys bought five^j roses. They^i took them^j home.

For the more Montagovian dynamic logics – DMG and DTT – a similar change, to state switchers that switch into multiple cases, would be needed.

There is a minor difference between the DS model and how dynamic logics associate an anaphor with an antecedent. Dynamic logics choose an antecedent and use the same variable. The DS model equates the information in one slot with the information in another. A small change to dynamic logics would make the same approach possible. For example,

(4) A man walks in the park. He whistles.

could receive a DPL translation of

(5) $\exists x[man(x) \wedge walk(x)] \wedge \exists y[x = y \wedge whistle(y)]$

In fact, since both variables will remain bound at the right hand end of this formula, we could also use

(6) $\exists x[man(x) \wedge walk(x)] \wedge \exists y[whistle(y)] \wedge x = y$

which may more convenient to derive compositionally: the translations of the sentences are generated and conjoined, and then the anaphor resolution procedure adds terms such as $x = y$ to the right hand end of the formula.

Dynamic logics tend to be described in total terms, with a single assignment function (or a set of them in DPLP) containing the information about all variables. Accessibility can be represented by a structural definition on formulae of the logic, or by an assignment function being undefined on the relevant variable. The latter case can, for singular data at least, be seen as related to semantic number agreement. Given an anaphor whose translation is the variable x, if the antecedent is inaccessible then the assignment function applied to x could yield an empty entity, which fails the number agreement condition of the anaphor. Assignment functions tend, in most formulations of dynamic logics, to be seen in extensional terms, with all bindings explicitly enumerated.

Discourse representation theory

In singular DRT (Kamp, 1981), the approach is similar to that of DPL. For quantification and conditionals, each possible context, containing an individual in each slot, is constructed and used to interpret the anaphors. Discourse referents label slots much as variables do in DPL.

With plurals (Kamp and Reyle, 1990), the situation is a little more complex. There are several sorts of plural reference. The two principal ones can be illustrated by

(7) Three^i men lifted a^j table.

On the collective reading of (7), a *single* context is constructed, in which slot i contains a sum of the three men, and slot j the table they jointly lifted. Continuing with

(8) They^i carried it^j upstairs.

also on a collective reading, the single context is used, with the (slots for the) anaphors equated to the (slots for the) antecedents. On the other hand, if (7) is given a distributive reading, *multiple* contexts are set up, one for each man and the table he lifted. To give a collective reading to (9) in this case,

(9) They^i carried them^j upstairs.

the abstraction operator (Σ) is applied to the DRS which defines each of the multiple contexts, in effect gathering them up into a single context with plural entities in each slot. Finally, to give a distributive reading to (9), the distribution operator ($\in$) is used to extract the individuals from the plural entity derived from *three men*. The DRS conditions from (7) are copied into the DRS for (9), and the individual discourse referent used to select the table for each man. In summary, DRT uses multiple contexts for distributive readings, and a single context for collective ones, with operators in DRS conditions and construction rules on DRSs to move between the two representations.

In choosing an antecedent, DRT does something similar to equating slots, by introducing a new discourse referent for anaphors and equating it with an existing one. As with dynamic logics, anaphor resolution can be incorporated by selecting different discourse referents from the ones available. It does not allow restructuring of the information: when there is an anaphor, then not only the discourse referent but all the conditions relating it to other discourse referents are respected. As with DPLP, no readings of (3) other than the one that respects the dependence relation of the antecedents are possible.

Summary

The properties of the theories are as follows:

TAI: single context, total and extensional representation. Semantic accessibility.

DPL/DPLP: multiple contexts, usually seen as total and extensional. Structural accessibility.

DRT: multiple, partial, non-extensional representation of contexts, together with gathering operations on contexts. Accessibility is primarily structural with some semantic constraints.

An interesting suggestion that follows is that DRT may be looked at from a DS perspective. Sub-DRSs can be seen as consisting of a universe of discourse referents and a partial DS which has slots labelled by just these discourse referents. The contents and partitioning of the DS are defined by the conditions of the sub-DRS. Copying and other manipulations of DRS conditions become relations between the DSs of the respective sub-DRSs, and accessibility becomes a matter of constraints, such as semantic number agreement, on these relations.

Similarly, we could reformulate the logic of TAI in non-extensional terms, where (as already suggested) the constraints on DSs are left unevaluated. More usefully, it should be possible to construct a logic in which the empty entity is not needed, and the complications it introduces are avoided. Roughly we would want

(10) Three^i boys bought five^j roses.

to give a DS which is defined only on slots i and j, and which is merged with DSs defining other slots as the discourse is built up. The merging process is also needed at the sub-sentential level: from

(11) Three^i boys who buy five^j roses own four^k vases.

a DS of slots i and j is merged with one of slots i and k.

5.1.2 Heim's claims

In section 3.1.4, four properties of DRT and File Change Semantics suggested by Heim were listed, namely:

1. Indefinites are not quantifiers, existential or otherwise.
2. Anaphoric pronouns are treated as bound variables, rather than definite descriptions.
3. Quantifying determiners and the conditional operator can bind multiple variables.
4. Variables not bound by a quantifying determiner or the conditional operator are given an existential interpretation.

The properties contrast with previous work in the semantics of anaphora, where:

1. Indefinites are existential quantifiers.
2. Anaphoric pronouns are semantically equivalent to (possibly complex) definite descriptions.
3. Quantifying determiners, frequency adverbs, and the hidden operator of generality in conditionals bind just one variable.
4. There is no need for default existential quantification of free variables.

TAI is much closer to the second set of claims. Indefinites are usually existential quantifiers, except as modified by the Variant Translation Principle. Quantifiers bind a single variable. Finally, anaphoric pronouns receive their interpretation from DS slots, which are defined by the semantic equivalent of a definite description of their antecedent, via the *Max* operator. The main exception to this point is the antecedents with numerals on a specific, at least reading; in this case, there is some indeterminacy in the content of the DS slot.

5.1.3 Accessibility and acceptability

DRT and dynamic logics take a structural approach to accessibility. In contrast, the TAI account is based on semantic number agreement, an approach which has received little attention in the past. As we have seen, there are advantages and disadvantages to both. DRT with plurals needs annotations on variables (*pl* marks) and some additional semantic constraints in the form of atomicity conditions, and even then runs into problems on examples which involve contingent acceptability, such as the "director/child/teacher" sentences of section 2.4. TAI requires the PVTP, which is a structural principle.

TAI models the acceptability of a sentence in semantic terms: a sentence is acceptable, for the purposes of anaphora, if a DS can be constructed for it. There has been little detailed discussion in the literature of what acceptability means and how it should be modelled. For example, in DRT, if it is found that there is no accessible antecedent for an anaphor, then the construction algorithm blocks. As Chierchia and Rooth (1984) have discussed, the unacceptability can be moved into the interpretation process, by making embedding functions be undefined on inaccessible discourse referents (see section 2.4). Similar remarks can be made about DPL. TAI moves the point where acceptability is recognised even further down the process. The translation procedure and the interpretation functions of the logic will not block unacceptable anaphors: a meta-language expression is always derived, even if it can never be satisfied. I do not think that there is a clear way of deciding whether this is better than an account which is primarily structural without a clearer statement of what acceptability means. The contingently acceptable examples already quoted show that some unacceptable sentences can only be recognised at the level of the model theoretic interpretation. The fact that acceptability can often be recognised without knowing the details of the situation being described seems, at first sight, to argue against the treatment used in TAI. However, it is possible to recognise non-contingent unacceptability in TAI, in a model-independent fashion. For example,

(12) Three^i cats stood in the street. ?It^i sang.

can be deduced to be unacceptable because the properties of the translation of *three* and of *it* place constraints on the DS which cannot be satisfied simultaneously, regardless of the details of the model. However, this is not to say that recognising the unacceptability in this way is the most convenient way in any particular application of TAI.

5.1.4 Classification of anaphora

In section 2.1, a system of classification for anaphors was presented. A clear distinction was made between bound variable anaphora and E-type pronouns. DRT and (by and large) dynamic logics do away with this division, treating them in a unified way. The same is true of L(GQA), but there is one point that is interesting. Sentence (13) ((85) in section 4.3) contains a bound variable anaphor and has truth condition (13.T).

(13) Every^i cat washes itself^i_i.

(13.T) $every'(cat^*, {}^*\{X \in \mathcal{D} \mid W\backslash(X,i)\Downarrow i \in wash^*(X)\}$

which, by the definition of *every'* and *the'*, is equivalent to

$$cat' \subseteq \{X \in \mathcal{E} \mid W\backslash(X,i)\Downarrow i \in wash'(X)\}$$

The definitions of restriction and extraction are such that $W\backslash(X,i)\Downarrow i$ is identically equal to X. Thus, in effect, the bound variable anaphor follows naturally from the L(GQA) account of anaphors as a whole.

5.1.5 Cataphora

Cataphora in DRT and TAI is straightforward. In DRT, the construction algorithm makes forward references possible, since it detaches any left-to-right processing of the discourse from the order in which DRS conditions are evaluated. TAI can also handle cataphora, since the notion of sequencing is not built into the translation procedure or the interpretation of the logic. For example, the first clause of (14)

(14) Ever since her^i childhood, Dorit^i has been extremely lazy.

could be taken as allowing any DS in which slot i contains one individual, and the second as rejecting any DSs where that individual is not Dorit. Dynamic logics – particularly the ones which take dynamic processing below sentence level – force the order of evaluation of a formula to be the same as the discourse that led to it, making a treatment of cataphora harder.

5.2 Internal evaluation of TAI

5.2.1 Flexibility and choosing a reading

Perhaps the most important difference between the TAI and DRT or dynamic logics is the degree to which they make alternative readings possible. In TAI, there is some degree of flexibility, while both DRT and DPL tend to force one reading. Kamp himself has indicated that this lack of flexibility is a shortcoming of DRT:

> It has become increasingly clear over the past ten years that the analysis of donkey anaphora which [File Change Semantics] and DRT put forward in the early eighties does not really fit the linguistic data as we have come to perceive them. In fact, these data seem to put into jeopardy the principal feature – the treatment of indefinites as free variables – which set those proposals from earlier accounts of natural language quantification and binding. As a consequence it now appears desirable to look for alternative explanatory frameworks. (Kamp, 1991b, p.177)

Current work such as Chierchia's DTT (on which Kamp was commenting) can be seen as a revival of the E-type account. Previous attempts to formalise an E-type account, notably that of Cooper (1979) suffered from some problems. Firstly, they made use of predicates that defined the referent of an anaphor, but without being explicit about how such predicates should be derived. Cooper's translation of the standard quantified donkey sentence is

$$\forall u[(farmer(u) \wedge \exists v[donkey(v) \wedge own(u,v)]) \rightarrow \exists x[\forall y[[S(u)](y) \equiv y = x] \wedge beat(u,x)]]$$

where S is a relation between farmers and the referent of the pronoun, but the semantics is inexplicit on how S is to be obtained. Furthermore, the reading may be as over-constrained as DRT's is if, for example, $S(u)$ is defined as "the donkey owned by u". Kamp comments:

> The E-type account *does* predict the uniqueness effect of (3) [most farmers who own a donkey beat it] if it stipulates that the description which interprets the pronoun is to be constructed explicitly from linguistic material in the restrictor of the quantifier *most.* (Kamp, 1991b, p.179)

The approach taken in TAI can be seen as similar to Cooper's E-type approach, but with an explicit statement of what S is. It is this explicitness which leads to the flexibility of readings: by using different translations of the determiner *a*, the different readings of the donkey sentence are made possible. However, there is some danger that this procedure, provided by the variant translation principle, may be too flexible. To continue with Kamp's commentary:

> If the claim that a certain pronoun is an E-type pronoun amounts to no more than that it is interpreted by some description, recovered by some syntactic, semantics and/or pragmatic strategy, then there isn't a lot that we can conclude. (Kamp, 1991b, p.179)

I have already suggested that properties of other determiners in the sentence may partially guide the variant translation principle. However, something more is needed; the conditions as given do nothing to explain why a unique antecedent reading is needed for Partee's daughter sentence, but a universal reading is strongly preferred for Heim's sage plant sentence, and why neither necessarily stands out for the standard quantified donkey sentence. The preference rules are a step towards this, but they lack a systematic account of how and when they should apply.

There are several other aspects of TAI about which similar points can be made, namely:

1. When are different input and output indices used in anaphor translations?

2. When do the quantifier scope rules apply?

3. When do the existence presupposition rules apply, for disjunction and negation?

4. When is the indefinite lazy reading made available?

5. When are non-intersective readings of determiners to be used?

Each of these was required to obtain the correct readings of certain sentences. As with the VTP, it may be possible to find principles to guide them. For example, in the case of negation and disjunction, Groenendijk and Stokhof (1991a) introduce an ambiguity between static and dynamic versions of the operators, one permitting anaphora and the other blocking it. In effect, this corresponds to forbidding the existence presupposition and allowing it. Groenendijk and Stokhof suggest that monotonicity may guide the choice of which translation is more appropriate (p.30-31). Roughly, each step in the discourse is required to be an upward monotone quantifier over states. Dynamic negation, which permits anaphora, reverses the monotonicity of the expression is applies to. Hence it should only be used when the expression is downward monotone, for example, if it is already negated. However, two of their own examples suggest it may not be so simple:

(15) It is not the case that John doesn't own a car. It is red and parked in front of his house.

(16) It is not the case that John doesn't own a book. ?It is on his desk.

Opinions about these sentences vary, but I find it much easier to accept (15) than (16), unless *a book* can be taken as referring to a specific book, a reading which is far from natural. It appears necessary to make some appeal to factors beyond the properties of determiners and logical connectives. Something similar might carry back to the single negation cases. Examples like

(17) John doesn't own a car. It is parked outside.

can again be read if there is a specific car in question, but it takes a more explicit statement of the context, and a rather unnatural reading of the first sentence to force it.

Another perspective comes from the study of alternative quantifier scopings, where issues of selecting and ordering readings have received detailed consideration. Moran and Pereira (1992) list a number of constraints and preference rules. Many of them are based on the presence of specific lexical items, or on certain structural configurations. However, they also note that pragmatic principles provide some guidance. In

(18) John visited every house on a street.

the favoured reading has *a street* with wide scope. Since houses are normally expected to be on a street, the narrow scope reading would contribute little information, while the wide scope reading tells us that each house is on the same street, and so is more informative. As an example of how this kind of idea could be applied to one of the other source of ambiguity, consider again (15)–(16). It is (or might be) a general property of books that one individual can be expected to own many, and that, to some degree, they are indistinguishable. The same claim is not in general true of cars. Hence, it is much easier to make the existence presupposition required to interpret (15). The further information that "it is red" also adds weight to this: the presupposed entity is specific enough that some of its properties can be enumerated.

A similar approach – looking beyond the literal interpretation to what is being communicated – can (tentatively) be employed to guide application of the variant translation principle. For example, suppose a speaker can be assumed to be following the Gricean maxim of Quantity, which Levinson (1983, p.101) states as:

> (i) make your contribution as informative as is required for the current purposes of the exchange.
> (ii) do not make your contribution more informative than is required.

Then out of the possible readings of Heim's sage plant sentence, the hearer might reasonably assume, in the absence of other information, that the speaker meant to communicate the universal reading, since otherwise they would be identifying one of the sage plants, and hence being more informative than was required. A related idea lies behind Kadmon's way of getting the universal reading from a semantics predicts the unique antecedent one: to assume that if there is no way of identifying the object to be considered as the unique antecedent, then all the possibilities are considered. Again, in

(19) Every man who has a daughter thinks she is the most beautiful girl in the world.

we would not allow the variant translation principle to apply, since the result, being a contradiction, could be seen as violating a Quality maxim.

For the selection of indices, the preference rule suggested that the input and output indices should be the same, so that the reading of

(3) Three^i boys bought five^j roses. They^i took them^j home.

is the "prominent" one, with each boy taking home the roses he bought. The less prominent reading is, generally, only to be taken when there is some other information to direct it, such as a contradiction arising from the prominent reading. Again, the Quantity maxim could be said to be violated in the less prominent case, since the hearer must construct extra information about the antecedents.

In section 4.4.4, I suggested that pragmatic properties might be used to explain why the indefinite lazy reading is sometimes preferred and sometimes not. Each of the other readings makes the referent of the pronoun identifiable without neeeding a choice from the hearer. In contrast, to give an indefinite lazy reading to, say, *every farmer who owns a donkey beats it*, the hearer must either choose some particular subset of the donkeys owned by each farmer, or must accept an indeterminate reading of the sentence where the exact referent is not identifiable. If there is no information to make the choice available in the former case, then the maxim of Quantity is violated, causing the indefinite lazy reading to be dispreferred. On the other hand, if it does not matter whether a farmer beats all of his donkeys, just one, or some – perhaps for reasons of topic – any reading other than the indefinite lazy one violates this same maxim.

I do not intend this description as any more than a first approximation to an account of how pragmatics should interact with the semantics, and nor do I mean to imply a commitment to Grice's theory. Arriving at a more complete identification and formalisation of the actual principles is one of the areas where research in this field could profitably be continued. The main implication for the current work is that the semantics must be flexible enough to allow all the possible readings, either by means of vagueness or through ambiguity.

5.2.2 The representation of context

The evaluation of TAI presented so far is largely concerned with how it relates to other theories. Two questions can also be raised about an internal issue, namely the representation of the context:

1. Does the context contain sufficient information to make all the empirical predictions required for a theory of noun phrase anaphora?
2. Does the context contain too much information, rendering it intractable to work with?

I think the answer to the first question is that there is enough information, for the restricted range of anaphora covered by the theory, but modification will be needed to cover some other classes of anaphora. In section 2.5.4, examples such as (20) were quoted, which rely on modal or temporal information for their interpretation.

(20) This year the president is a Republican. Next year he will be a Democrat.

A possible solution is to pair either the whole DS or individual members of it with world and time indices. Nor are DSs rich enough to cope with paycheque sentences, for example

(21) The man who gave his paycheque to his wife was wiser than the one who gave it to his mistress.

For *it* in (21), a translation as a function from the subject antecedent (*the man*) to the referent of the anaphor would suffice, which could appear in the context either extensionally or in some unevaluated representation. The latter comes close to a purely syntactic approach, where the literal text of the antecedent is copied in place of the anaphor.

The question of whether there is too much information is a harder one. Most of the complexity of DSs comes from the need to represent dependence relations. From

(22) Three boys bought five roses.

the problem might not be too bad. The number of ways of partitioning each antecedent is fairly small, and if the exact relations are not known, it may be feasible to enumerate all the plausible possibilities. But with the example that lead Scha to introduce the cumulative reading,

(23) 500 Dutch firms use 6000 American computers.

the size of the problem gets out of hand. As we have seen, things become even worse where more than two antecedents are involved, and where the technical device of the empty entity $\perp$ is needed. Note that this problem need not arise if the truth conditions are all that are needed: since the extensions are closed under summation, all that matters is that the right number of objects can be found, and the essential relations between them are unimportant.

A solution is to move to an account which leaves the dependence relations between the antecedents vague. There are two ways of doing this. Either we drop the idea of representing dependence explicitly, so that slots only have content and no structure. Or, the constraints which specify the dependence between antecedents are retained, but are left unevaluated except where there is something which indicates that they need to be evaluated. The main argument against dropping dependence altogether is the fact that readings of anaphors which respect the dependence relations are more prominent than those that do not for "small" examples such as (22). Adopting the second approach defers the problem to wider questions of language processing (from the linguistic perspective) or of how the semantics is to be used (from a computational one). In particular, some reference would be needed to the level of detail intended by the speaker of an utterance and required by the hearer of it, as might happen if a hearer questioned whether

(3) Three^i boys bought five^j roses. They^i took them^j home.

was true, and demanded the details of who did what.

5.3 Extensions to TAI

The current benchmark for theories of anaphora is the Kamp and Reyle (1990) version of DRT. No other theories have as wide a coverage of the data. Kamp and Reyle discuss a number of phenomena concerning plural noun phrases which I will omit: floating quantifiers, dependent plurals, reciprocals and *same*, *different* and *together*. Many of the insights and principles discussed by Kamp and Reyle can be incorporated into TAI, in the same way that it was possible to parallel the account of the logical connectives. I will briefly mention three possible extensions to the coverage of TAI: partial full NPs, anaphora under restriction, and combined antecedents. The answers I propose should be taken as first sketches, rather than complete solutions.

5.3.1 Partial full NP anaphora

Partial full NP anaphora is exemplified by

(24) Some animals are standing in a field. The^i_j sheep are grazing.

The sheep is anaphoric to *some animals* but does not take all of its content. Partial full NP anaphors can be added to TAI by essentially the same trick used for indefinite lazy readings in section 4.4.4. The standard full NP anaphor translation is followed, with the substitution of *some* for *the*. That is, the definite determiner is translated as

$$\lambda s_1.\lambda i.\lambda s_2.(some\ Of(j, P))(i)\ (s_1 \bigwedge s_2)$$

which takes the input slot j and requires that some part of it is in $s_1 \bigwedge s_2$; this part is left in the output slot, i. The second sentence in (24) receives the translation

(24a) $(some\ Of(j, pl))(i)\ (sheep_s \bigwedge graze_s)$

and the interpretations

(24.T) $some'(\sharp(W{\downarrow}j), sheep^* \cap graze^*)$

(24.A) $W{\Downarrow}i \in Max(\sharp(W{\downarrow}j) \cap sheep^* \cap graze^*) \wedge pl'(W{\Downarrow}i)$

Paraphrasing, there is some part of slot j which consists of sheep that graze, and this part of the slot forms the content of the output slot i. A disadvantage of this account is that it does not allow the structure of the input to be kept in the output, as we might want in (25).

(25) Some animals each belong to a (different) farmer. The donkeys are beaten by them.

meaning each donkey is beaten by the farmer it belongs to. A way of overcoming this, albeit an inelegant one, is to add an expression to the formula which has the interpretation $\forall X \in W{\downarrow}i[W\backslash(X, i) = W\backslash(X, j)]$; that is, for each thing in the output slot i, all dependence relations of the input slot j are respected in slot i.

This is only one way of looking at partial full NP anaphora. Another is to say that there is a mechanism which creates new antecedents, and hence new slots in the DS, behind the scenes. There is independent motivation for such a mechanism, namely the examples of section 2.5.1 which involved making bridging assumptions, such as:

(26) Patience walked into a room. The chandeliers burned brightly.

for which the bridging assumption was

(26a) The room referred to by *a room* had chandeliers in it.

In the case of (24), the bridging assumption would be:

(27) The sheep referred to by *the sheep* are part of the animals referred to be *some animals*.

which can quite naturally be derived from the background knowledge that sheep are animals.

5.3.2 Anaphora under restriction

In section 2.4, the need for semantic agreement under restriction was noted for two constructs: partitives, as in (28) and (29), and the floating quantifier *each*, as in (30).

(28) Every child got one present and two sweets.
One of them unwrapped the present/?the presents and ate the sweets/?the sweet.

(29) Every director gave a present to a child from the orphanage. One of them opened it/its present straight away.

(30) Every farmer owns a donkey. They each beat it.

There is a parallel between the floating quantifier and the partitives, in that the second sentence of (30) can be rephrased as *each of them beat it.*

The logic already contains a term which provides the right semantics, namely $Of_r(j, P)$, which extracts information from the DS under restrictors. The semantic agreement predicate for (28)–(30) is derived from the syntactic number of the anaphor: *sg* if singular, and *pl* if plural. For floating quantifiers, another translation principle needs to be added, which forces the variant translation of the pronoun when *each* appears in the VP containing it. A sketch of how this might be done in syntactic terms runs as follows: have a feature on the pronoun which indicates that the variant translation is to apply, and unify it with a corresponding feature on the VP node. The feature defaults to a non-restricted reading using *Of*. If *each* is present, it changes the feature on the VP node, and Of_r is used.[2]

For the partitive case, the same approach can be used, with the feature appearing on the partitive construction. An additional construct in the logic for the partitive itself is also needed. One possibility is to translate *one of them* as something like $(1_{\geq}\ Of(i, un))$ which is interpreted as the generalized quantifier

$$\{X \subseteq \mathcal{D} \mid \exists x \in (X \cap \sharp(W{\downarrow}i))[Crd(x) \geq 1]\}$$

i.e. the set of sets containing at least one member from slot i. This gives the partitive a reading of "at least one of them". The other quantifiers for numerals are possible. Perhaps the most likely one is the specific, at least interpretation, $(\Phi\ 1(Of(i, un)))$.

5.3.3 Combined antecedents.

In section 4.2.4.2, the problem of "combined antecedents" was raised, illustrated by (31) and (32).

(31) John^i met Mary^j after work. They went for a walk.

(32) John^i owns a^j dog and Mary^k owns a cat^l. They got them from the pet shop.

In (31), the i and j slots have to summed to form an antecedent for *they.* (32) shows that any dependence relations of the i and k slots must also be merged. There is already a mechanism in L(GQA) for the formal operation of combining the slots, namely the *Merge* operator of section 4.5.3. What remains to be found is some means of deciding when to form a new antecedent from two others. If computational effort is not an issue, then

[2] It may be better to analyse *each* as being part of the NP, in which case the feature could be unified between *each* and the NP node, and thence with the VP node.

all combinations of antecedents could be formed. In practice, however, we will generally want to combine the antecedents only if there is some reason to do so. The same happens in DRT: any two or more accessible discourse referents may be summed, but Kamp and Reyle (1990) suggest that the summation principle be applied only "when a suitable antecedent must be created for some anaphoric noun phrase (p.738)".

It may be possible to use properties of the lexical items and syntactic relations to guide the process. For example, there is a parallel sentence to (31) in which *met* is used symmetrically, and the antecedents are conjoined:

(31a) John^i and Mary^j met after work. They went for a walk.

In the case of (32), a parallel between the conjuncts is established because they have the same verb, and this might be used to trigger combining the subject slot from one phrase with that of the other, and likewise the objects. However, there are cases where there are no obvious principles to help:

(33) John gave Mary a present. They were on holiday at the time.

(34) John was hiding in the coal shed. Mary was disguised as a tree. They were trying to avoid Fred.

In cases like this, we have to look to more general criteria, perhaps drawing on the same information used in the selection of plausible antecedents for an anaphor. Thus in (33), the anaphor requires an animate antecedent, and hence "John and Mary" is a possible antecedent, but "John and the present" is not. Similarly, it may be possible to detect some connection between the first two sentences of (34): the causal background for both comes from the third sentence.

5.4 Computational Aspects of TAI

Most of the discussion so far has been concerned with the linguistic properties of TAI. In this section I look at two aspects of TAI from the perspective of computational linguistics. The first is how TAI relates to Bonnie Webber's work on anaphora, which has some similarities with E-type theories. Secondly, I look at how truth conditions and DSs may be computed from L(GQA) formulae. Although L(GQA) has not yet been implemented, the algorithm I sketch suggests that it is possible to do so in an efficient way.

5.4.1 Webber's theory

The theory of anaphora developed by Webber (1979, 1981, 1983) starts from the idea of *discourse model synthesis*, according to which one objective of discourse is to communicate a model from a speaker to a hearer, and the process of "understanding" an utterance is equated with the synthesis of an appropriate model by the listener. Her theory includes "one" anaphora and VP ellipsis as well as definite (noun phrase) anaphora; I will concentrate on the latter. There are five assumptions behind her work, stated in Webber (1981):

1. One objective of discourse is to enable a speaker to communicate to a listener a model s/he has of some situation.

2. Such a discourse model can be viewed as a structured collection of entities.

3. A speaker uses a definite anaphor to refer to an entity in his/her discourse model (DM_S), on the assumptions that there will be a similar model in the listener's discourse model (DM_L), and that the listener will be able to access and identify that entity via the definite anaphor.

4. The referent of a definite anaphor is an entity in DM_S which the speaker presumes to have a counterpart in DM_L.

5. In deciding which discourse entity a definite anaphor refers to, a listener's judgements stem in part from how the entities in DM_L are described.

What Webber means by her fifth point is that the manner in which the speaker evokes the referent is important as well as the referent invoked. Consider (35) and (36):

(35) John took two trips around France. They were both wonderful.

(36) John travelled around France twice. ?They were both wonderful.

In (35), the first sentence explicitly evokes the two trips, making them available in the discourse model. In (36), the existence of the two trips can be inferred, but the trips themselves are not evoked, and cannot be referred to by an anaphor. This idea is a well-known one in anaphora: it is essentially the same point made by Partee's "marbles" examples quoted in section 2.5.2:

(37) I lost ten marbles and found all but one of them. It is probably under the sofa.

(38) I lost ten marbles and found only nine of them. ?It is probably under the sofa.

There are similarities between Webber's assumptions and those of both DRT and TAI. In Kamp's view of DRT as a theory of verbal communication discussed in section 3.1.1, the process of communication is identified with the construction in the hearer's mind of a DRS resembling the speaker's one, which roughly corresponds to Webber's first, second and fourth assumptions. Webber's fifth point is (in effect) addressed in DRT by making the DRS construction algorithm work only from the surface form of the sentence: no additional inference is used in forming discourse referents. Very similar remarks apply to TAI, with the synthesis of a discourse model corresponding to the process of constructing a context in the form of a DS or set of possible DSs, and the constraints on DSs being derived from the surface form by a compositional procedure.

Webber formalises her approach by stating a number of rules for computing *initial descriptions* (IDs) for discourse entities. IDs are expressions which provide enough information to identify the referent of the antecedent. Anaphors are translated into IDs, which take the form of definite descriptions. Webber speaks of an antecedent *evoking* the discourse entity and the anaphor *accessing* it. The ID rules are stated as syntactic operations on the logical form of the sentences, and apply after quantifier scoping. She does not give a model-theoretic interpretation to the logical forms. Most of the time this does not matter, but there is one place where, as we shall see, it leaves certain important properties of the formalism vague. The procedure for translating natural language sentences into formulae is also not described. A standard compositional procedure is sufficient.

Noun phrases are represented as restricted quantifiers of the form $(Qx : P)$, where P is a predicate derived from the nominal. Quantifiers are applied to open formulae. For example

(39) Some boy is happy.

becomes

$$(\exists x : Boy).Happy(x)$$

Predicates may be basic, formed by lambda abstraction from a formula, or constructed using a special operator *set*. Lambda abstraction is used to restrict quantifiers, so that the NP

(40) A peanut that Wendy gave to a gorilla

is represented as

$$\exists x : \lambda(u : Peanut)[(\exists y : Gorilla).Gave(Wendy, u, y)]$$

which reads as "an x which is a u such that u is a peanut and there is a gorilla y such that Wendy gave u to y". The operator *set* takes a predicate P as argument, and forms a predicate which is true of sets of things that satisfy P. This appears to lead to a type mismatch, which can easily be overcome by identifying non-set objects with singleton sets. The logic also includes a set membership operator, used with universal quantification to represent distributive uses of plurals. Thus

(41) Each man ate a pizza

has the representation

$$(\exists! w : set(Man))(\forall x \in w)(\exists y : Pizza)Ate(x, y)$$

read as "there is a unique w which is a set of men, such that for each member x of w, there is a pizza which x ate".

Webber's rules start with a formula consisting of a sequence of quantifiers applied to an open formula. A discourse entity is constructed from the leftmost quantifier. The formula is then rewritten with the quantifier removed and a variable standing for the discourse entity substituted for the variable over which the quantification took place. There have been a number of different statements of the rules; to illustrate how they work, the rule for indefinite NPs follows, as it is presented in Webber (1983):

Rule 1. The formulae on the left produce the IDs on the right.

$(\exists X : Q).P(X)$	$\iota X : Q(X) \wedge P(X) \wedge Evoke(S, X)$
$(\exists X : set(Q)).P(X)$	$\iota X : set(Q)(X) \wedge P(X) \wedge Evoke(S, X)$
$(\exists X : set(Q))(\forall x \in X).P(x)$	$\iota X : set(Q)(X) \wedge (\forall x \in X)P(x) \wedge Evoke(S, X)$

ι is the Russellian definite operator, meaning "the (unique) object such that ...". S is a label identifying the sentence and clause containing the antecedent. The meaning of *Evoke* is explored below; for now it can be thought of as just a way of recording where the object came from. Taking one of Webber's examples,

(42) I saw a cat
$(\exists x : Cat).Saw(I, x)$

the corresponding ID is

$$\iota X : (Cat(X) \wedge Saw(I, X) \wedge Evoke(S_{42}, X))$$

That is, "the cat I saw that was evoked by sentence (42)".

In Webber (1983), there are also rules for quantifiers formed from definite NPs, and for iterated contexts arising from distributive quantification. For example, in

(43) Each cat ate a mouse it saw.
$(\exists! w : Cat)(\forall y \in w)(\exists x : \lambda(u : Mouse)[Saw(y, u)]).Ate(y, x)$

"each cat" is first processed. A discourse entity e_1 is created for the set of cats, and the formula rewritten to:

$$(\forall y \in e_1)(\exists x : \lambda(u : Mouse)[Saw(y, u)]).Ate(y, x)$$

which then produces the ID:

$$\{x \mid (\exists y \in e_1) Mouse(x) \wedge Ate(y, x) \wedge Evoke(S_{43}, x)\}$$

meaning the set of things each of which is a mouse and for each of which there is a cat who saw it and ate it and which was evoked by (43).

Donkey sentences are analysed as follows. The representation of

(44) Every farmer who owns a donkey beats it.

is, following the example in Webber (1981):

$$(\forall x : \lambda(u : Farmer)[(\exists y : Donkey).Own(u, y)]).Beat(x, IT)$$

where the pronoun *IT* is to be replaced by the ID derived from *a donkey*. From the indefinite NP rule, this is:

$$\iota y : Donkey(y) \wedge Own(x, y) \wedge Evoke(S_{44.1}, y)$$

where $S_{44.1}$ means the clause containing *a donkey*. The ID says that y is the donkey which x owns and which was evoked by $S_{44.1}$. Substituting into the full formula,

$$(\forall x : \lambda(u : Farmer)[(\exists y : Donkey).Own(u, y)]).$$
$$Beat(x, \iota y : Donkey(y) \wedge Own(x, y) \wedge Evoke(S_{44.1}, y))$$

This is a unique anaphor reading: from the semantics of the ι operator, just one donkey is evoked per farmer. Webber gives some idea of how the universal reading, which she takes to be the correct one, may be obtained:

> I noted earlier that the "evoke" term in an ID is a preliminary attempt to link an entity in the listener's discourse model with its presumed counterpart in the speaker's model. By itself, it does not commit either the speaker or the listener to believing that there is only one such x that has the other properties given in its ID, or that the entity corresponds to something that exists in the real world. (Webber, 1981, p.293)

Although she does not make the semantics of *Evoke* any more precise than this statement, presumably it means that $Evoke(S_{44.1}, y)$ in the above formula is to hold if y is the set of all donkeys for a given farmer, and not just a specific one. There is a further rule which produces an ID for intersentential reference, as in

(45) Every farmer who owns a donkey beats it. They are planning to get back at them.

The ID for the whole collection of donkeys is

$$\{w \mid (\exists x : \lambda(u : Farmer)[(\exists y : Donkey).Own(u, y)]). \\ w = \iota z : Donkey(z) \wedge Own(x, z) \wedge Evoke(S_{45}, z)\}$$

The exact membership of this collection again depends on the interpretation of *Evoke*.

This mechanism provides the flexibility that the data of section 2.3 suggested was necessary. For example, to get the indefinite lazy reading, *Evoke* could require that y is just some of the donkeys per farmer. What is lacking is a systematic statement of what factors affect the interpretation of *Evoke*, a criticism very similar to that raised against Cooper's formalisation of the Evans analysis. For example, with

(46) Every man who has a daughter will leave her all his vintage port.

Webber acknowledges that the universal reading is not possible, and suggests that "world knowledge" will

> remind one that whereas a man who owns a donkey can beat each one of the donkeys he owns, a man who has a daughter cannot leave each one *all* his vintage port. (Webber, 1981, p.293)

However, the details of what form this "world knowledge" might take, and how it is to be applied are left unstated. There is a danger in allowing it to be too flexible. For example, the example (36) (*John travelled around France twice. ?They were both wonderful.*), which Webber herself indicated was unacceptable, would be permitted if the world knowledge permitted the inference that whenever someone travels around a country, a trip which can be referred to is evoked.

Webber's work is widely cited in the computational literature. An implementation of it is described by Ayuso (1989), who claims it to be the first such implementation. Ayuso modifies Webber's formalism in a number of ways, including the introduction of intensional contexts. *Evoke* is replaced by a Skolem function, and a pointer outside the logical form to the NP in the parse tree. The result is that the discourse entity evoked by an indefinite NP is a single individual, so that donkey sentences would receive a unique anaphor reading. Ayuso mentions that "trickier cases" of donkey sentences, and interactions with negation, have not yet been addressed.

5.4.1.1 Webber and TAI

Webber's work has received almost no attention in the theoretical linguistics work on anaphora, which – even given the problems with *Evoke* – is perhaps surprising. The most complete version of her theory was published as Webber (1979), at about the same time as Evans' principal paper on the subject (Evans, 1977) and the related work by Cooper (1979), and it has something in common with their work. In particular, Webber's evoked discourse referents are very much like the definite descriptions which fix the referent of E-type pronouns. As discussed in chapters 2 and 3, "E-type" theories were soon superseded by the "indefinites as variables" account of Kamp (1981) and Heim (1982). TAI is closer to the E-type account than to the Kamp-Heim one, and it may therefore be interesting to look at the parallels between it and Webber's work.

The relation between the theories is easiest understood by considering an example. The sentence

(47) Some farmers (each) own a donkey.

appears in Webber's formalism as:

$$(\exists X : set(Farmer))(\forall x \in X)(\exists y : Donkey).Own(x, y)$$

For subsequent reference by

(48) They beat them.

the discourse referents accessed are e_1 and e_2:

$$e_1 : \iota X : set(Farmer)(X) \wedge (\forall x \in X)(\exists y : Donkey).Own(x, y) \wedge Evoke(S_{47}, X)$$
$$e_2 : \{y \mid Donkey(y) \wedge (\exists x \in e_1).Own(x, y) \wedge Evoke(S_{47}, y)\}$$

meaning e_1 is the set made up of farmers who each own a donkey, and e_2 is the set of donkeys such that there is a member of e_1 who owns the donkey. An L(GQA) translation is

$$(some\ farmer_s)(i)\ \hat{x}[(1_{\geq}\ donkey_s)(j)\ \hat{y}[own(x, y)]]$$

for which the anaphoric part of the interpretation is:

$$W\Downarrow i \in Max(farmer^* \cap {}^*\{X \in \mathcal{D} \mid 1'_{\geq}(donkey^*, own^*(X))\}) \wedge$$
$$W\downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W\backslash(X, i)\Downarrow j \in Max(donkey^* \cap own^*(X)) \wedge$$
$$W\backslash(X, i)\downarrow j \subseteq own_\perp{}'(X)\}$$

Slot i is essentially the same as e_1, and given that the condition $\exists x \in e_1$ is analogous to saying that X comes from $W\downarrow i$, slot j is essentially the same as e_2. There are two major differences between the interpretations: the structuring in L(GQA), and the use of *Evoke* by Webber. Structuring is missing in Webber's approach: her representation of (48) will (presumably) be

$$Beat(THEY, THEM)$$

which becomes

$$Beat(\iota X : set(Farmer)(X) \wedge (\forall x \in X)(\exists y : Donkey).Own(x, y) \wedge Evoke(S_{47}, X),$$
$$\{y \mid Donkey(y) \wedge (\exists x \in e_1).Own(x, y) \wedge Evoke(S_{47}, y)\})$$

This says that the entire collection of entities e_1 beats the entire collection of entities e_2. Representing it as if it were *they each beat them* will not help either: the result is a reading where each farmer beats all of the donkeys. However, it would probably be possible to add another rule which would take apart e_2 into its members, by matching against $\exists x \in e_1$, and replacing it with the individual variable bound to the members of e_1 by *They each....* Another approach might be to modify the *Evoke* predicate, so that it allowed extra arguments, in this case the variable for the individual members of *they*, and found the corresponding donkey. The usual problem of defining the semantics of *Evoke* precisely arises on this solution. The other role of *Evoke*, in identifying the constituent that gave rise to a discourse entity is captured in the TAI mechanism of assigning each NP an index.

TAI, while in some ways following similar ideas to Webber, has improved over it in two principal ways. Firstly, there is an explicit representation of dependence, allowing the distinct readings of (48). Secondly, properties of *Evoke* which can be stated (more or less) systematically have been separated out. For example, in Webber's approach the universal reading of donkey sentences requires *Evoke* to generalise from one donkey owned by a farmer to all of them, while in TAI it is captured by the variant translation principle, following from the properties of *every*. As we have noted, deciding when the variant translation principle is appropriate, and when other semantic and pragmatic principles must be brought to bear is not a trivial issue. However, by identifying some of the possible principles, we have gained over Webber's appeal to "world knowledge".

5.4.2 Implementation techniques

In this section, I present a means of computing with L(GQA) expressions in an efficient way. The method is illustrated with a reduced version of L(GQA) which excludes the logical connectives and *Of* terms. The results generalise to the full version. The idea is to introduce a notational device which lifts out all non-trivial expressions occurring more than once, and to design a representation for sets which means that the structuring information is computed with no additional effort.

To lift out common sub-expressions, a more formal version of the "where ..." notation used for the examples of section 4.3 is employed. Any meta-language expression may be followed by *where* $f = P$, meaning that P is to be computed and the result substituted wherever f occurs in the expression. f may take parameters, and *where* expressions may be nested, with inner ones inheriting parameters from ones enclosing them. For example

$$P \textit{ where } f(X) = (Q \textit{ where } g = R)$$

means that occurrences of X in R are bound by f, even though X does not occur as a parameter of g. In the examples below, the implicit parameters are expanded out, so that the above would become

$$P \textit{ where } f(X) = (Q \textit{ where } g(X) = R)$$

The truth conditional and anaphoric parts of the interpretation function are combined into a single function $[\![.]\!]^{W,m}$, the result of which is a pair. As usual, $\downarrow$ is used to extract the components. The only difference from previous use of tuples is that the components may be sets.

The interpretation function for the reduced version of L(GQA) is defined as follows:

$$[\![Q(i)S]\!]^{W,m} = ([\![Q]\!]^{W,m}i)([\![S]\!]^{W,m}i)$$

$$[\![R(x_1,\ldots,x_n)]\!]^{W,m} = \langle R^*(m_1(x_1),\ldots,m_1(x_n)), R_\perp{}'(m_1(x_1),\ldots,m_1(x_n))\rangle$$

$$[\![DS]\!]^{W,m} = \lambda i.\lambda q.([\![D]\!]^{W,m}i)([\![S]\!]^{W,m}i)$$

$$[\![\hat{x}[F]]\!]^{W,m} = \lambda i.\langle {}^*\{X \in \mathcal{D} \mid f(X)\downarrow 1\}, res(W,m)\downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid f(X)\downarrow 2\}\rangle$$
$$\textit{where } f(X) = [\![F]\!]^{W,m[x/\langle X,i\rangle]}$$

$$[\![R_s]\!]^{W,m} = \lambda i.\langle R^*, \textit{true}\rangle$$

$$[\![D]\!]^{W,m} = \lambda i\lambda p.\lambda q.\langle D'(p\downarrow 1, f), res(W,m)\Downarrow i \in Max_E(f) \wedge p\downarrow 2 \wedge q\downarrow 2\rangle$$
$$\textit{where } f = p\downarrow 1 \cap q\downarrow 1$$

if D is a distributive determiner

$$
\begin{aligned}
[\![D]\!]^{W,m} &= \lambda i \lambda p.\lambda q.\langle D'(p{\downarrow}1, f), res(W,m){\Downarrow}i \in Max(f) \wedge p{\downarrow}2 \wedge q{\downarrow}2\rangle \\
& \quad \textit{where } f = p{\downarrow}1 \cap q{\downarrow}1 \\
& \quad \text{otherwise}
\end{aligned}
$$

There are a few minor differences from the earlier statement of the interpretation functions. The clause for determiners now takes an extra parameter q, and the interpretation functions contain $D'(p,q)$ directly rather than $q \in \{X \subseteq D \mid D'(p,X)\}$. Secondly, the conservativity property of determiners is exploited to replace $D'(p,q)$ by $D'(p, p \cap q)$.

Taking an earlier example,

(49) Two^i boys buy three^j roses.

(49a) $(2_{=}\ boy_s)(i)\ \hat{x}[(3_{=}\ rose_s)(j)\ \hat{y}[buy(x,y)]]$

the new interpretation is:

(49b) $\langle 2'_{=}(boy^*, p), W{\Downarrow}i \in Max(p) \wedge W{\downarrow}i \subseteq \{X \in \mathcal{D}_\perp \mid q(X){\downarrow}2\}\rangle$
$\quad$ *where* $p = boy^* \cap {}^*\{X \in \mathcal{D} \mid q(X){\downarrow}1\}$
$\quad$ *where* $q(X) = \langle 3'_{=}(rose^*, r(X)),$
$\qquad W\backslash(X,i){\Downarrow}j \in Max(r(X)) \wedge W\backslash(X,i){\downarrow}j \subseteq \{Y \in \mathcal{D}_\perp \mid s(X,Y){\downarrow}2\}\rangle$
$\quad$ *where* $r(X) = rose^* \cap {}^*\{Y \in \mathcal{D} \mid s(X,Y){\downarrow}1\}$
$\quad$ *where* $s(X,Y) = \langle buy^*(X,Y), buy_\perp{}'(X,Y)\rangle$

As before, we will also need a top-level function which forms the conjunction of the truth conditional and anaphoric parts as the overall interpretation of a closed formula.

The more important part of making L(GQA) computationally efficient concerns the representation of sets. The basic idea is that any member of a derived set (i.e. something of the form $\{X \mid \ldots\}$) is annotated with a reason for why the member is present in the set. The reason makes reference only to predicates of two or more places, and identifies the subset of the *essential* extension of the predicate that contributed to the presence of the member, and which component of the tuples was relevant. For example, suppose we have the set

$$^*\{Y \in \mathcal{D} \mid buy^*(X,Y)\}$$

with $X = x_1 \oplus x_2$, where x_1 buys y_1 and x_2 buys (separately) y_2 and y_3, so that

$$buy' = \{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle, \langle x_2, y_3\rangle\}$$

Then the set will contain the following members and annotations:

$y_1 \oplus y_2$ because $\{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle\} \subseteq buy'$

$y_1 \oplus y_3$ because $\{\langle x_1, y_1\rangle, \langle x_2, y_3\rangle\} \subseteq buy'$

$y_1 \oplus y_2 \oplus y_3$ because $\{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle, \langle x_2, y_3\rangle\} \subseteq buy'$

In each case, the annotation should also say that it was the second component of *buy'* that was relevant. In the set of X's of which $x_1 \oplus x_2$ is a member, the annotation must record that the first component was relevant, and that any of the three subsets of *buy'* are possible reasons for its inclusion. The information about which subset of *buy'* should be placed in the annotation can be recorded with each member of *buy**.

To see how this helps, take an expression such as

$$W \Downarrow i \in Max(p) \wedge W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid \ldots\}$$

In the worked examples in section 4.3, the content of slot i ($W \Downarrow i$) was evaluated first, and then each sub-sum of the content considered as a possible member of $W \downarrow i$, some of them subsequently being rejected by the structuring conditions. By the *Max* condition, $W \Downarrow i$ must be a member of p, the argument of *Max*. Ultimately, slot i will be structured by a predicate P which appears in p. Hence, by recording which members of the essential extension of P were summed to form $W \Downarrow i$, we can assume that the relevant part of $W \downarrow i$ is made up of exactly those members, and test only them for membership in $\{X \ldots\}$. In the above example, we might have that $W \Downarrow i = x_1 \oplus x_2$, but in the condition $W \downarrow i \subseteq \{X \ldots\}$, only $X = x_1$ and $X = x_2$ need be tried, since the remaining sub-sum of $W \Downarrow i$, namely $x_1 \oplus x_2$, does not appear in the annotations derived from the essential extension *buy'*. The structuring on slot j is similarly derived from the annotations, meaning that the final term, $buy_\perp{}'(X, Y)$ need not be evaluated at all.

Some other minor optimisations can also be applied: when slot i is seen to be $\perp$, all slots which depend on it are also $\perp$. When the annotations on members of slot i are empty, as would happen from

(50) Some^i farmers buy no^j donkey.

or from verb phrase negation, the j slot can be set to be empty, and any arbitrary structuring of slot i used. Note also that the procedure finds the minimal DS.

The approach can be illustrated with the above example and the model used in section 4.3.2.5, in which we had:

$$boy' = \{b_1, b_2, b_3\}$$

$$rose' = \{r_1, r_2, r_3, r_4\}$$

$$buy' = \{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle, \langle c, r_4\rangle\}$$

Assume that the evaluation of (49b) proceeds by starting with the truth conditions, and once these are fully evaluated, the anaphoric conditions are examined. Furthermore, when we have expressions such as $boy^* \cap {}^*\{X \in \mathcal{D} \mid \ldots\}$, each possible sum from boy^* is broken down into its sub-sums, and the results substituted for X in turn, following from the sum closure operations. In this example, the determiner can also be used to simplify the process, by only considering sums of size 2 in boy^*.

With the given model, the evaluation goes as follows. The sums of size 2 in boy^* are $b_1 \oplus b_2$, $b_1 \oplus b_3$ and $b_2 \oplus b_3$. Each of these sums must be tested for membership in ${}^*\{X \in \mathcal{D} \mid q(X){\downarrow}1\}$, which in turn entails testing each sum or any sub-sum of it for membership in $\{X \in \mathcal{D} \mid q(X){\downarrow}1\}$. $q(X)$ is also evaluated, which in turn evaluates $r(X)$ and hence $s(X, Y)$ for Y equal to each rose or sum of roses. For the time being, only the first components of the pairs $q(X)$ and $s(X, Y)$ are needed. Evaluating $s(X, Y)$ simply means looking at members of the extension buy^*. In doing so, any X containing b_3, since it does not appear in the buy relation. At this stage each X and Y value is annotated with a subset of *buy'*. The following values then satisfy $s(X, Y)$, with their annotations:

$X = b_1$, $Y = r_1$ because $\{\langle b_1, r_1\rangle\} \subseteq buy'$

$X = b_2$, $Y = r_2 \oplus r_3$ because $\{\langle b_2, r_2 \oplus r_3\rangle\} \subseteq buy'$

$X = b_1 \oplus b_2$, $Y = r_1 \oplus r_2 \oplus r_3$ because $\{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus r_3\rangle\} \subseteq buy'$

After evaluating $3'_=(\ldots)$, only the last of these is still possible, since the others do not contain three roses in Y. Hence p can contain one member, which is (with its annotation):

$b_1 \oplus b_2 \in p$ because $\{\langle b_1, r_1\rangle, \langle b_2, r_2 \oplus b_3\rangle\} \subseteq buy'$

This satisfies the condition from $2'_=(\ldots)$, showing the formula to be true in the model.

Evaluation can now proceed to the anaphoric part. The condition $W \Downarrow i \in Max(p)$ immediately tells us that slot i contains $b_1 \oplus b_2$, since this is the maximal (and only) member of p. To evaluate

$$W \downarrow i \subseteq \{X \in \mathcal{D} \mid q(X) \downarrow 2\}$$

we look to the annotations, which tell us that $W \downarrow i$ was formed from a subset of buy' which contained b_1 and b_2, but not $b_1 \oplus b_2$. Hence, the latter member need not be tested for membership in $\{X \in \mathcal{D} \mid q(X) \downarrow 2\}$. Setting $X = b_1$, the evaluation of the second component of $q(X)$ first requires finding $Max(r(X))$. $r(X)$ is already known to be r_1 with its annotation as above, and so $W \backslash (b_1, i) \Downarrow j = r_1$. For this member, there is only one way it could be structured, but for $X = b_2$, we have $W \backslash (b_2, i) \Downarrow j = r_2 \oplus r_3$. Looking at the annotation tells us that the DS has to contain $\langle b_2, r_2 \oplus r_3\rangle$ and not $\langle b_2, r_2 \oplus r_3\rangle$ and $\langle b_2, r_3\rangle$. Finally, the structuring condition follows trivially.

The above description is a sketch of how to evaluate L(GQA) expressions than a complete account, and some work would be needed to turn it into a computable algorithm. For example, a sentence with three antecedents such as

(51) Two^i boys who buy three^j roses own four^k vases.

has an anaphoric interpretation of the form:

$$\begin{array}{l} W \Downarrow i \in Max(boy^* \cap {}^*\{X \in \mathcal{D} \mid 3'_=(rose^*, buy^*(X))\} \cap \\ \quad {}^*\{X \in \mathcal{D} \mid 4'_=(vase^*, own^*(X))\}) \wedge \\ W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W \backslash (X, i) \Downarrow j \in Max(rose^* \cap buy^*(X)) \wedge \ldots\} \wedge \\ W \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid W \backslash (X, i) \Downarrow k \in Max(vase^* \cap own^*(X)) \wedge \ldots\} \end{array}$$

There are two expressions of the form $W \downarrow i \subseteq \{X \ldots\}$. In evaluating one, the algorithm must know to take the partitioning of slot i which is defined by the extension buy', and for the other as defined by own'. This can be deduced by the algorithm, by a simple examination of the body of the term $\{X \ldots\}$, or by adding a suitable annotation to $W \downarrow i$ in the interpretation.

The overall conclusion is that DSs can be evaluated with almost no computational effort over that needed for the truth conditions. This suggests that a computationally efficient implementation of L(GQA) in an extensional framework is possible.

5.5 Conclusions

5.5.1 Summary

My motivation in developing a new theory of anaphora has come from two sources. The empirical evidence of chapter 2 showed the interpretation of sentences involving anaphora depends on a wide range of factors, even for sentences which are broadly similar in syntactic and semantic structure. In the case of donkey sentences, at least the following make a difference to the reading: the particular determiners which are present, properties of the situation being described (daughter and sage plant sentences), and the context provided by other parts of the discourse. Intuitions about such sentences also vary from one language user to another. This means that a theory of the semantics of anaphora has to be flexible enough to provide a number of readings. This could be achieved by making sentences ambiguous between the possibilities, by assigning a single reading which is weaker than all of them, or ideally, by generating all plausible readings with an order of preference on them. Principles which predict the order of preference should be guided by the factors mentioned above.

The major current theories of anaphora – DRT and dynamic logics – tend to force one reading, typically the universal one, and this reading is deeply embedded in the formalism. TAI has attempted to rectify this by providing a logic which obtains the different readings by making certain determiners ambiguous, at the level of the translation procedure. Selection between readings is realised in preference rules, applied to devices such as the variant translation principle, which relates the reading of indefinite noun phrases to properties of other noun phrases in a sentence.

The second strand to the research is to question the methodological approach taken in DRT and dynamic logics. Both theories adopt unconventional semantic representations, in which both model-theoretic and structural properties are important in assigning the meaning to an expression. In DRT, indefinites gain their meaning from the discourse representation structure in which they appear, the form of which is dictated by the surface form of the discourse without reference to the truth conditions. For dynamic logics, the order that terms appear in formulae affects the interpretation they receive. Despite its apparent complexity, the logic used in TAI is relatively conventional, being static and not relying on the representation in the way that DRT does. All aspects of the interpretation are expressed in model-theoretic terms. A goal of adopting such a logic (though as yet unrealised) is to integrate it with work in other areas of semantics, and to be able to use existing computational technology. DRT and dynamic logics also rely on structural definitions of acceptability. I have tried to question whether other definitions are possible, by developing a theory of acceptability which is primarily semantic, with the conclusion that such an approach will work provided a small amount of additional structural information is made available. A final methodological difference from the existing theories is that I have tried to make the underlying conceptual model clearer: to state what information appears to be necessary to interpret anaphora, independently of how that information is represented and extracted. Although the model I propose is expressed in terms which are closely related to its formalisation in L(GQA), I think it nevertheless provides a way of understanding how the three theories are related to one another, independently of the fine details of the formalisms they employ.

5.5.2 Directions for further research

The principles used in the translation procedure to make different readings available and to indicate preferred ones are a first attempt. In section 5.2.1, I suggested that some reference must be made to what a speaker wishes to communicate, which is consistent with the variability with context and between language users for donkey sentences. DRT has touched on the issues of verbal communication and cognitive representation, but not in a way that has had an obvious empirical influence on its theory of anaphora. Two other factors that may prove to be important are the topic and focus structure, which was seen in section 2.3.2 to influence the reading of conditional donkey sentences, and temporal/event information, in the same way that Verkuyl and van der Does (1991) have suggested might disambiguate distributive from collective readings (section 2.6). Further research in this area could usefully concentrate on first refining the data, with the aim of identifying what factors are important and how they interact, and then arriving at a formalisation of how these factors affect the sentence reading. It seems likely that some of the more concrete statements of pragmatic principles will be of value here, such as those of Grice.

Where it has been possible to state a principle which guides the selection of one reading over another, it would be valuable to look for generalisations relating the principle to other semantic properties. For example, the variant translation principle leads to a universal reading of donkey sentences as the preferred one when the subject determiner is *every*. In section 2.3.1, it was noted that *no* also favours the universal reading. What these determiners have in common is that they make a claim about all objects in the domain, and we might search for other related properties, to guide the degree to which the related preference rule should apply.

A related question concerns the status of preference rules and of rules such as the VTP and the widening rules used for quantifier scoping and disjunction. Specifically, it is important to ask what they should operate on: the syntactic representation, L(GQA) expressions or meta-language expressions? Although the empirical predictions will be the same from a semantic point of view, there may be differences from other perspectives, for example when performance properties are considered, psycholinguistically, philosophically, or computationally.

The view of what the semantics is used for has been a very narrow one. It (by and large) assumes that a discourse is processed from its surface form to its meta-language equivalent, and that the truth and anaphoric conditions are then evaluated with respect to a model to yield a truth value and its contribution to the context. This is rarely what utterances are used for, and it would be interesting to see how the logic might be used for other (and more realistic) communicative acts. An example is *wh*-questions, such as

(52) Which farmers own a donkey?

which could be interpreted by taking a translation equivalent to *some farmers own a donkey*, and reporting the content of the relevant DS slot. In doing so, the further information about the donkeys each farmer owns can also be made available, for questions such as *which ones?* Similarly, generation is of interest, in effect running the process the other way. In the context of anaphora, this could mean, for example, establishing what dependence relation is required between anaphors, and then finding an utterance which is consistent with the corresponding DS.

Finally, I have suggested that the relatively conventional logic should make it possible to integrate TAI with other pieces of semantic theory, and to apply conventional computational technology to it. Section 5.4.2 went some way towards demonstrating the latter point, but both issues need further exploration. The behaviour of anaphors in modal and temporal contexts might provide a testing ground for the former, as might the (possibly related) issue of anaphora in relation to attitude verbs explored by Asher (1987). Integration with Montagovian or situational frameworks would be of interest in both cases.

A Appendix: DRT in detail

This appendix presents DRS languages for singular and plural anaphora, in the style of Kamp and Reyle (1990). See section 3.1 for a description of DRT.

A.1 DRT for singular anaphora

Language

The DRS language used by Kamp for his treatment of singular anaphora is as follows:

1. Vocabulary
   (a) a set V of discourse referents.
   (b) a set N of proper nouns.
   (c) sets $R_1, R_2, \ldots$ of $1-, 2-, \ldots$ place predicate symbols.
2. DRSs and DRS conditions
   (a) if U if a (finite) subset of V and *Con* a (finite) set of DRS conditions, then $\langle U, Con\rangle$ is a DRS. U is called the universe of K.
   (b) if $x, y \in Con$, $x = y$ is a DRS condition.
   (c) if R is an n-place predicate symbol, and $x_1, \ldots, x_n \in V$, then $R(x_1, \ldots, x_n)$ is a DRS condition.
   (d) if K_1 and K_2 are DRSs, then $K_1 \Rightarrow K_2$ is a DRS condition (called an *implicative* condition).

Syntactic fragment

Kamp treats (approximately) the following fragment of English, recast as a phrase structure grammar.

1. Syntactic categories and lexicon
   (a) PRO (Pronoun): *he*, *she*, *it*, *him*, *her*
   (b) PN (Proper noun): *Pedro*, *Chiquita*, *John*, *Mary*, ...
   (c) N (Common noun): *farmer*, *donkey*, ...
   (d) VP (Verb phrase/intransitive verb): *thrives*, *walks*, ...
   (e) V (Transitive verb): *owns*, *beats*, ...
   (f) S (Sentence)
2. Formation rules
   (a) S $\rightarrow$ NP VP
   (b) S $\rightarrow$ *if* S, S
   (c) VP $\rightarrow$ V NP, provided the noun phrase is not *he* or *she*
   (d) NP $\rightarrow$ *a(n)* N
   (e) NP $\rightarrow$ *every* N
   (f) NP $\rightarrow$ PN
   (g) NP $\rightarrow$ PRO
   (h) N $\rightarrow$ N *who* VP

Construction rules

The rules are presented in the form of a triggering configuration, and one or more actions. The triggering configuration is a parse tree, given here as a labelled bracketing. Where a node of the parse tree is irrelevant to the rest of the rule, only the category is given; otherwise a syntactic variable and its category are given. The actions introduce new conditions and discourse referents, and rewrite part or all of the triggering configuration. K is the DRS being processed, $U.K$ its universe and $Con.K$ its conditions. Actions may apply to either K or the topmost DRS, K_0. The rules have the same coverage as Kamp's, but are stated slightly differently. I have omitted a few minor details to do with gender and case marking. The relative clause rule RC is also stated slightly differently from Kamp and Reyle's NRC rule.

PN Triggered by $[[[\alpha]_{PN}]_{NP}\ VP]_S$ or $[V\ [[\alpha]_{PN}]_{NP}]_{VP}$.
Enter a new discourse referent u into $U.K_0$ and $\alpha(u)$ into $Con.K_0$. Replace $[[\alpha]_{PN}]_{NP}$ by u.

PRO Triggered by $[[[\alpha]_{PRO}]_{NP}\ VP]_S$ or $[V\ [[\alpha]_{PRO}]_{NP}]_{VP}$.
Choose a suitable and accessible discourse referent v. Enter a new discourse referent u into $U.K$ and $u = v$ into $Con.K$. Replace $[[\alpha]_{PRO}]_{NP}$ by u.

ID Triggered by $[[a(n)\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [\text{a(n)}\ [\alpha]_N]_{NP}]_{VP}$.
Enter a new discourse referent u into $U.K$ and $[\alpha]_{N(u)}$ into $Con.K$.
Replace $[\text{a(n)}\ [\alpha]_N]_{NP}$ by u.

LIN Triggered by $[\alpha]_{N(u)}$. Replace triggering configuration by $\alpha(u)$.

RC Triggered by $[[\alpha]_N\ \text{who}\ [\beta]_{VP}]_{N(u)}$.
Enter $\alpha(u)$ into $Con.K$. Replace triggering configuration by $[u[\beta]_{VP}]_S$.

COND Triggered by $[\text{if}\ [\alpha]_S,\ [\beta]_S]_S$.
Replace triggering configuration by $K_1 \Rightarrow K_2$, where K_1 and K_2 are empty DRSs. Enter $[\alpha]_S$ into $Con.K_1$ and $[\beta]_S$ into $Con.K_2$.

EVERY Triggered by $[[\text{every}\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [\text{every}\ [\alpha]_N]_{NP}]_{VP}$.
Replace triggering configuration by $K_1 \Rightarrow K_2$, where K_1 and K_2 are empty DRSs. Enter a new discourse referent u into $U.K_1$, $[\alpha]_{N(u)}$ into $Con.K_1$ and γ into $Con.K_2$, where γ is the result of substituting u for $[\text{every}\ [\alpha]_N]_{NP}$ in the triggering configuration.

There is no rule for verbs or verb phrases; the rules for noun phrases eventually rewrite the lexical form of the NP as a discourse referent, leaving conditions such as $x\ [\alpha]_{VP}$ and $x\ [\alpha]_V\ y$, which may be taken as equivalent to the conditions $\alpha(x)$ and $\alpha(x, y)$, respectively. Kamp and Reyle do not make this final step, and instead give direct verification conditions for $x\ [\alpha]_{VP}$ and $x\ [\alpha]_V\ y$.

Accessibility

A discourse referent u is accessible from a DRS K if:

1. $u \in U.K$; or
2. K appears in a condition of the form $K_1 \Rightarrow K$, and $u \in U.K_1$; or
3. K is a sub-DRS of K_1, and u is accessible from K_1.

Truth definition

Assume a model $M = \langle E, F \rangle$, where E is a non-empty set of individuals, and F a function which assigns to each proper noun an element of E, and to each n-place predicate symbol a set of n-tuples of members of E. F maps each proper noun to its unique bearer in E. A DRS K is true in M if there is an embedding function f which verifies K in M. Embedding functions have discourse referents as their domain and E as their range. An extension of an embedding function f to a set of discourse referents D is an embedding function g which has domain equal to the union of D and the domain of f, and which is such that for all discourse referents u in the domain of f, $f(u) = g(u)$. We then say that g extends f to D, written $f \subseteq_D g$. Verification is defined by:

1. f verifies K in M iff f verifies each of the conditions of K.

2. f verifies a condition γ in M, written $M \models_f \gamma$, as follows:

   (a) $M \models_f x = y$ if $f(x) = f(y)$
   (b) $M \models_f R(x_1, \ldots, x_n)$ if $\langle f(x_1), \ldots, f(x_n) \rangle \in F(R)$
   (c) $M \models_f K_1 \Rightarrow K_2$ if for every $f \subseteq_{U.K_1} g$ such that $M \models_g K_1$, there is a $g \subseteq_{U.K_2} h$ such that $M \models_h K_2$.

A.2 DRT for plural anaphora

Additional conditions:

1. $at(x)$ and $non\!-\!at(x)$.

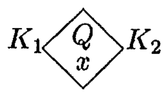
2. $x = y_1 \oplus \ldots \oplus y_n$.

3. $x = \Sigma z K$, where z is a discourse referent in $U.K$.

4. $|x| = \nu$, where ν is $1, 2, 3$, etc.

5. $x \in y$

6. Duplex conditions, $K_1[Qx]K_2$, where Q is a determiner which can act as a quantifier, and x is a discourse referent in the universe of K_1. Kamp and Reyle use the notation:



   $K_1[Qx]K_2$ is a little more convenient typographically.

Plural discourse referents X are ones accompanied by the condition $non\!-\!at(X)$; singular ones x have the condition $at(x)$. Discourse referents with neither condition are called neutral discourse referents.

Construction rules

The construction rules of the previous section are modified, so that PRO applies only to singular antecedents, and the following rules are added:

NP.def Triggered by $[[[\delta]_{Det}\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [[\delta]_{Det}\ [\alpha]_N]_{NP}]_{VP}$, where δ is a definite plural determiner.
Enter a new plural discourse referent U into $U.K_0$, and $[\alpha]_{N(U)}$ into $Con.K_0$. Replace $[[\delta]_{Det}\ [\alpha]_N]_{NP}$ by U.

NP.ind Triggered by $[[[\delta]_{Det}\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [[\delta]_{Det}\ [\alpha]_N]_{NP}]_{VP}$, where δ is an indefinite plural determiner.
Enter a new plural discourse referent U into $U.K$, and $[\alpha]_{N(U)}$ into $Con.K$. Replace $[[\delta]_{Det}\ [\alpha]_N]_{NP}$ by U.

Summation Triggered by a sub-DRS K' of the DRS K (or K itself), and two or more discourse referents $y_1, \ldots, y_n$ occurring in K and accessible from K'.
Enter a new plural discourse referent U into $U.K'$ and $Z = y_1 \oplus \ldots \oplus y_n$ into $Con.K'$.

Abstraction Triggered by a condition $K_1[Qx]K_2$ in $Con.K$.
Enter a new discourse referent Y into $U.K$ and $Y = \Sigma z K'$ into $Con.K$, where K' is a new DRS such that $U.K' = U.K_1 \cup U.K_2$ and $Con.K' = Con.K_1 \cup Con.K_2$, and z is a member of $U.K'$.

NP.quant Triggered by $[[[\delta]_{Det}\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [[\delta]_{Det}\ [\alpha]_N]_{NP}]_{VP}$, where δ is a quantifying determiner.
Replace triggering configuration by $K_1\ [\delta u]\ K_2$, where K_1 and K_2 are empty DRSs. Enter a new discourse referent u into $U.K_1$, $[\alpha]_{N(u)}$ into $Con.K_1$ and γ into $Con.K_2$, where γ is the result of substituting u for $[[\delta]_{Det}\ [\alpha]_N]_{NP}$ in the triggering configuration. If δ is a plural determiner, mark u with *pl*. This rule replaces EVERY.

PRO.plu Triggered by $[[\beta]_{PRO}\ VP]_S$ or $[V\ [\beta]_{PRO}]_{VP}$, where β is plural.
Choose a discourse referent Y which is accessible, and which is either plural, marked *pl* or marked *pl(u)* and the triggering configuration contains either u or a discourse referent marked *pl(u)*. Enter a plural discourse referent X into $U.K$ and $X = Y$ into $Con.K$. Replace $[\beta]_N$ by X.

Distribution (Optional rule). Triggered by $[X\ VP]_S$ or $[V\ X]_{VP}$.
Enter a condition K_1 [*every* x] K_2 into $Con.K$, where x is a new discourse referent, $U.K_1 = x^{pl}$, $Con.K_1 = x \in X$, $U.K_2$ is empty, and $Con.K_2$ contains a condition resulting from replacing X in the triggering condition by x. Delete triggering condition.

DA ("Distribution over abstraction") Given a DRS K, such that $U.K$ contains discourse referents U and Z, and $Con.K$ contains $Z = U$ (or a sequence $Z = Z_1, Z_1 = Z_2, \ldots, Z_n = U$), $Z = \Sigma z K_1$, and

u^{pl}	
$u \in U$	[*every* u] K_2

Replace the left hand side of the last of these conditions by K', such that $U.K'$ contains u^{pl}, and each discourse referent from $U.K_1$ except z, annotated with $pl(u)$, and $Con.K'$ contains $u \in U$ and the conditions from $Con.K_1$ with u substituted throughout for z.

Genera Triggered by a constituent β of category N.
Introduce a new discourse referent U into $U.K$ and a condition $\beta * (U)$ into $Con.K$, where $\beta*$ is a predicate that holds of both single objects and collections of objects.

Some rules, for example dependent plurals, have been omitted. PRO.plu is stated slightly differently from Kamp and Relye's formulation.

A gap in the rules as given in Kamp and Reyle (1990) is that noun phrases with an indefinite plural determiner introduce a discourse referent and a condition from the noun, but no condition on the size of the plural discourse referent. The rule should therefore be modified to something like:

NP.ind Triggered by $[[[\delta]_{Det}\ [\alpha]_N]_{NP}\ VP]_S$ or $[V\ [[\delta]_{Det}\ [\alpha]_N]_{NP}]_{VP}$, where δ is an indefinite plural determiner.
Enter a new plural discourse referent U into $U.K$, and $[\alpha]_{N(U)}$ and $\delta'(U)$ into $Con.K$, where δ' is the translation of the determiner. Replace $[[\delta]_{Det}\ [\alpha]_N]_{NP}$ by U.

Indefinite plural determiners include plural *some*, for which the translation, like singular *some*, introduces no extra conditions, and numerals, which add a condition of the form $|X| = \nu$.

Truth definition

The statement of the verification conditions is kept relatively informal here, to avoid bringing in the extra notation of Link's theory of plurals. f verifies a condition γ in M, written $M \models_f \gamma$, as follows:

1. $M \models_f at(x)$, if $f(x)$ is an atom of M.
2. $M \models_f non\!-\!at(x)$, if $f(x)$ is a non-atomic entity (sum) of M.
3. $M \models_f x = y_1 \oplus \ldots \oplus y_n$, if $f(x)$ is equal to the sum of $f(y_1), \ldots, f(y_n)$.
4. $M \models_f x = \Sigma z K$, if $f(x)$ is equal to the sum of the b for which $M \models_g K$, where $g = f \cup \{\langle z, b\rangle\}$.
5. $M \models_f |x| = \nu$, if $f(x)$ is made up from ν atoms.
6. $M \models_f x \in y$, if $f(x)$ is an atom of M, $f(y)$ is a non-atom of M, and $f(x)$ is contained in $f(y)$.
7. $M \models_f K_1[Qx]K_2$, if $\langle A, B\rangle \in R$, where Q denotes R, and
$A = \{b \mid \exists g[f \cup \{\langle x, b\rangle\} \subseteq_{U.K_1 - \{x\}} g \wedge M \models_g K_1]\}$
$B = \{b \mid \exists g[f \cup \{\langle x, b\rangle\} \subseteq_{U.K_1 - \{x\}} g \wedge M \models_g K_1] \wedge$
$\forall g[f \cup \{\langle x, b\rangle\} \subseteq_{U.K_1 - \{x\}} g \Rightarrow \exists h[g \subseteq_{U.K_2} h \wedge M \models_h K_2]]\}$

B Appendix: TAI rules

This appendix summarises the meta-language operators, the formation rules and interpretation functions of L(GQA) and L(GQAD) and the syntax rules of the English fragment.

B.1 Meta-language operators

Basic operators: standard set and truth operators, plus $\oplus, \downarrow, \langle . \rangle$.

Properties and definitions:

1. $\perp$: $\forall x \in \mathcal{E}_\perp[x \oplus \perp = x]$
2. $\oplus$: $a \oplus b = b \oplus a$; $(a \oplus b) \oplus c = a \oplus (b \oplus c)$; $a \oplus a = a$
3. $Ats(a)$ is the set of atoms A such that $\oplus A = a$. $Ats(\perp) = \emptyset$.
4. $Crd(a)$ is the sum-cardinality: $Crd(a) = card(Ats(a))$.
5. $Sup(A)$ is the $a \in A$ such that $\forall a' \in A[(a \neq a') \to Crd(a') < Crd(a)]$
6. $\langle a_1, \ldots, a_n \rangle \downarrow \langle i_1, \ldots, i_k \rangle = \langle a_{i_1}, \ldots, a_{i_k} \rangle$. Defined only if, for each i_j, $1 \leq i_j \leq n$.
7. $W \downarrow \langle i_1, \ldots, i_n \rangle = \{w \downarrow \langle i_1, \ldots, i_n \rangle \mid w \in W\}$
8. $W \Downarrow \langle i_1, \ldots, i_n \rangle = \oplus(W \downarrow \langle i_1, \ldots, i_n \rangle)$
9. $\oplus\{\langle a_1, \ldots, a_n \rangle, \ldots, \langle z_1, \ldots, z_n \rangle\} = \langle \oplus\{a_1, \ldots, z_1\}, \ldots, \oplus\{a_n, \ldots, z_n\} \rangle$
10. $W \backslash (X, i) = \{w \in W \mid w \downarrow i = X\}$
11. $^*S = \{Y \in \mathcal{D}^n \mid \exists X \subseteq S[Y = \oplus X]\}$
12. $\sharp(S) \; = \; ^*(\mathrm{Ats}(\oplus S))$
13. $S(x_1, \ldots, x_n) \; = \; \langle x_1, \ldots, x_n \rangle \in S$
14. $S_\perp(x_1, \ldots, x_n) \; = \; (\langle x_1, \ldots, x_n \rangle \in S) \vee (x_1 = \perp) \vee \ldots \vee (x_n = \perp)$
15. $Distr(S) = \forall s \in S[Ats(s) \subseteq S]$
16. $Max(S) = \begin{cases} \{s \in S \mid \forall t \in S[Crd(t) \leq Crd(s)]\} & \text{if } S \neq \emptyset \\ \{\perp\} & \text{otherwise} \end{cases}$
17. $Max_E(S) = \begin{cases} Max(S) & \text{if } Distr(S) \\ \emptyset & \text{otherwise} \end{cases}$
18. $V \downarrow i = \{W \downarrow i \mid W \in V\}$
19. $Merge(W, i, j, k) \; = \; (W \Downarrow i = W \Downarrow j \oplus W \Downarrow k) \wedge$
    $\forall X \in W \downarrow j[W \backslash (X, i) = \emptyset \vee W \backslash (X, i) = W \backslash (X, j)] \wedge$
    $\forall X \in W \downarrow k[W \backslash (X, i) = \emptyset \vee W \backslash (X, i) = W \backslash (X, k)]$
20. $\sum V = \{\oplus W \mid W \in V\}$

B.2 L(GQA)

Syntax

Basic expressions:

1. Determiners, symbolised as D below. Examples: *most*, *few*, *every*, *no*, *the*, Φ and for each integer number $n \geq 1$, the symbols $n_=$, $n_\geq$ and $n_\leq$.
2. Variables: $x, y, \ldots$.

3. Predicate symbols: R. Examples: *buy*, *walk*, *farmer*, *own*, *beat*, *john*, *mary*. Each predicate has a specified number of argument places.
4. Indices: i, j.
5. Number terms: n, where n is a natural number.

Derived expressions:

1. Quantifiers: Q.
2. Set terms: S.
3. Formulae: F.

Formation rules:

F1. If Q is a quantifier, i is an index and S is a set term, $Q(i)S$ is a formula.

F2. If R is an n-place predicate symbol, and $x_1, \ldots, x_n$ are variables, $R(x_1, \ldots, x_n)$ is a formula ($n \geq 2$).

F3. If D is a determiner and S a set term, DS is a quantifier.

F4. If x is a variable and F is a formula, $\hat{x}[F]$ is a set term.

F5. If R is a predicate symbol, R_s is a set term.

F6. If S_1 and S_2 are set terms, $S_1 \bigwedge S_2$ is a set term.

F7. If n in a natural number and S is a set term, $n(S)$ is a set term.

F8a. If j is an index and P is a 1-place predicate, $Of(j, P)$ is a set term.

F8b. If j is an index and P is a 1-place predicate, $Of_r(j, P)$ is a set term.

F9-or. If F_1 and F_2 are formulae, then $F_1 \bigvee F_2$ is a formula.

F9-and. If F_1 and F_2 are formulae, so is $F_1 \bigwedge F_2$.

F10-or. If R_1 and R_2 are n-place predicate symbols, then $R_1 \bigvee R_2$ is an n-place predicate.

F10-and. If R_1 and R_2 are n-place predicate symbols, then $R_1 \bigwedge R_2$ is an n-place predicate.

F11-or. If S_1 and S_2 are set terms, then $S_1 \bigvee S_2$ is a set term.

F12-or. If Q_1 and Q_2 are quantifiers and i_1 and i_2 are indices, then $Q_1(i_1) \bigvee Q_2(i_2)$ is a quantifier.

F12-and. If Q_1 and Q_2 are quantifiers and i_1 and i_2 are indices, then $Q_1(i_1) \bigwedge Q_2(i_2)$ is a quantifier.

F13. If F is a formula, then $\neg F$ is a formula.

F14. If S is a set term, then $\neg S$ is a set term.

Interpretation

Top-level interpretation:

$$[\![F]\!]^W = [\![F]\!]_t^{W,[\,]} \wedge [\![F]\!]_a^{W,[\,]}$$

Truth conditional part:

$[\![Q(i)S]\!]_t^{W,m} = [\![S]\!]_t^{W,m} i \in [\![Q]\!]_t^{W,m} i$

$[\![R(x_1, \ldots, x_n)]\!]_t^{W,m} = [\![R]\!]_t^{W,m}(m_1(x_1), \ldots, m_1(x_n))$

$[\![DS]\!]_t^{W,m} = \lambda i.([\![D]\!]_t^{W,m} i)([\![S]\!]_t^{W,m} i)$

$[\![\hat{x}[F]]\!]_t^{W,m} = \lambda i.^*\{X \in \mathcal{D} \mid [\![F]\!]_t^{W,m[x/\langle X,i\rangle]}\}$

$[\![R_s]\!]_t^{W,m} = \lambda i.[\![R]\!]_t^{W,m}$

$[\![S_1 \bigwedge S_2]\!]_t^{W,m} = \lambda i.([\![S_1]\!]_t^{W,m} i) \cap ([\![S_2]\!]_t^{W,m} i)$

$[\![n(S)]\!]_t^{W,m} = \lambda i.n' \cap ([\![S]\!]_t^{W,m} i)$

$[\![R]\!]_t^{W,m} = R^*$

$[\![D]\!]_t^{W,m} = \lambda i \lambda p.\{X \subseteq \mathcal{D} \mid D'(p, X)\}$, where D' is the determiner definition.

$[\![Of(j, P)]\!]_t^{W,m} = \lambda i.\sharp(W \downarrow j)$

$[\![Of_r(j, P)]\!]_t^{W,m} = \lambda i.\sharp(res(W, m) \downarrow j)$

$[\![F_1 \bigvee F_2]\!]_t^{W,m} = [\![F_1]\!]_t^{W,m} \vee [\![F_2]\!]_t^{W,m}$

$[\![R_1 \bigvee R_2]\!]_t^{W,m} = [\![R_1]\!]_t^{W,m} \cup [\![R_2]\!]_t^{W,m}$

$[\![S_1 \bigvee S_2]\!]_t^{W,m} = \lambda i.[\![S_1]\!]_t^{W,m} i \cup [\![S_2]\!]_t^{W,m} i$

$[\![Q_1(i_1) \bigvee Q_2(i_2)]\!]_t^{W,m} = \lambda i.([\![Q_1]\!]_t^{W,m} i_1) \cup ([\![Q_2]\!]_t^{W,m} i_2)$

$[\![F_1 \bigwedge F_2]\!]_t^{W,m} = [\![F_1]\!]_t^{W,m} \wedge [\![F_2]\!]_t^{W,m}$

$[\![R_1 \bigwedge R_2]\!]_t^{W,m} = [\![R_1]\!]_t^{W,m} \cap [\![R_2]\!]_t^{W,m}$

$[\![Q_1(i_1) \bigwedge Q_2(i_2)]\!]_t^{W,m} =$
$\lambda i.\{X \subseteq \mathcal{D} \mid \exists Y \in ([\![Q_1]\!]_t^{W,m} i_1) \cap ([\![Q_2]\!]_t^{W,m} i_2)[Sup(^*Y) \in X]\}$

$[\![\neg F]\!]_t^{W,m} = \sim [\![F]\!]_t^{W,m}$

$[\![\neg S]\!]_t^{W,m} = -[\![S]\!]_t^{W,m}$

Anaphoric part:

$[\![Q(i)S]\!]_a^{W,m} = [\![Q]\!]_a^{W,m} i([\![S]\!]_t^{W,m} i)([\![S]\!]_a^{W,m} i)$

$[\![R(x_1, \ldots, x_n)]\!]_a^{W,m} = [\![R]\!]_a^{W,m}(m_1(x_1), \ldots, m_1(x_n))$

$[\![DS]\!]_a^{W,m} = \lambda i \lambda t \lambda a.([\![D]\!]_a^{W,m})t([\![S]\!]_t^{W,m} i)a([\![S]\!]_a^{W,m} i)$

$[\![\hat{x}[F]]\!]_a^{W,m} = \lambda i.res(W, m) \downarrow i \subseteq \{X \in \mathcal{D}_\perp \mid [\![F]\!]_a^{W,m[x/\langle X,i\rangle]}\}$

$[\![R_s]\!]_a^{W,m} = \lambda i.true$

$[\![S_1 \bigwedge S_2]\!]_a^{W,m} = \lambda i.([\![S_1]\!]_a^{W,m} i) \wedge ([\![S_2]\!]_a^{W,m} i)$

$[\![n(S)]\!]_a^{W,m} = \lambda i.[\![S]\!]_a^{W,m} i$

$[\![R]\!]_a^{W,m} = R_\perp{}'$

$[\![D]\!]_a^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2.res(W, m) \Downarrow i \in Max_E(t_1 \cap t_2) \wedge a_1 \wedge a_2$
if D is a distributive determiner

$[\![D]\!]_a^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2.res(W, m) \Downarrow i \in Max(t_1 \cap t_2) \wedge a_1 \wedge a_2$ otherwise.

$[\![Of(j, P)]\!]_a^{W,m} = \lambda i.P'(W \Downarrow i)$

$[\![Of_r(j, P)]\!]_a^{W,m} = \lambda i.P'(W \Downarrow i)$

$[\![F_1 \bigvee F_2]\!]_a^{W,m} = [\![F_1]\!]_a^{W,m} \vee [\![F_2]\!]_a^{W,m}$

$[\![R_1 \bigvee R_2]\!]_a^{W,m} = [\![R_1]\!]_a^{W,m} \vee [\![R_2]\!]_a^{W,m}$

$[\![S_1 \bigvee S_2]\!]_a^{W,m} = \lambda i.[\![S_1]\!]_a^{W,m} i \vee [\![S_2]\!]_a^{W,m} i$

$$
\begin{aligned}
[\![Q_1(i_1) \bigvee Q_2(i_2)]\!]_a{}^{W,m} &= \lambda i \lambda t \lambda a.(([\![Q_1]\!]_a{}^{W,m} i_1\ t\ true) \vee ([\![Q_2]\!]_a{}^{W,m} i_2\ t\ true)) \\
&\qquad \wedge a \wedge Merge(W, i, i_1, i_2) \\
[\![F_1 \bigwedge F_2]\!]_a{}^{W,m} &= [\![F_1]\!]_a{}^{W,m} \wedge [\![F_2]\!]_a{}^{W,m} \\
[\![R_1 \bigwedge R_2]\!]_a{}^{W,m} &= [\![R_1]\!]_a{}^{W,m} \wedge [\![R_2]\!]_a{}^{W,m} \\
[\![Q_1(i_1) \bigwedge Q_2(i_2)]\!]_a{}^{W,m} &= \lambda i \lambda t \lambda a.([\![Q_1]\!]_a{}^{W,m} i_1\ t'\ true) \wedge ([\![Q_2]\!]_a{}^{W,m} i_2\ t'\ true) \wedge \\
&\qquad W \Downarrow i \in t \wedge a \wedge Merge(W, i, i_1, i_2) \\
&\text{where } t' = \{X \in \mathcal{D} \mid \exists Y \in t[Ats(X) \subseteq Ats(Y)]\} \\
[\![\neg F]\!]_a{}^{W,m} &= [\![F]\!]_a{}^{W,m} \\
[\![\neg S]\!]_a{}^{W,m} &= [\![S]\!]_a{}^{W,m}
\end{aligned}
$$

Auxiliary function:

$$res(W, m[x/\langle X, i\rangle]) = res(W \backslash (X, i), m)$$
$$res(W, [\,]) = W$$

Determiner definitions, $D'(p, q)$:

Distributive determiners, applied to $card(p \cap \mathcal{E})$ and $card(p \cap q \cap \mathcal{E})$.

every: $\lambda mn.m = n$

most: $\lambda mn.n \geq m/2$

few: $\lambda mn.n \leq f(m)$, for some measure of "fewness" f.

some: $\lambda mn.n \geq 1$

no: $\lambda mn.n = 0$

Other determiners and alternative definitions.

$$
\begin{aligned}
some'(p, q) &= p \cap q \neq \emptyset \\
no'(p, q) &= p \cap q = \emptyset \\
all'(p, q) &= Sup(p) \in q \\
n'_{\geq}(p, q) &= \exists x \in p \cap q[Crd(x) \geq n] \\
n'_{\leq}(p, q) &= \exists x \in p \cap q[Crd(x) \leq n] \\
n'_{=}(p, q) &= \exists x \in Max(p \cap q)[Crd(x) = n] \\
\Phi'(p, q) &= p \cap q \neq \emptyset \\
the'(p, q) &= Sup(p) \in q
\end{aligned}
$$

Modification for non-intersective determiners

$[\![D]\!]_a{}^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2 . res(W, m) \Downarrow i \in Max_E(t_1) \wedge a_1$ if D is a distributive determiner

$[\![D]\!]_a{}^{W,m} = \lambda i \lambda t_1 \lambda a_1 \lambda t_2 \lambda a_2 . res(W, m) \Downarrow i \in Max(t_1) \wedge a_1$ otherwise.

B.3 L(GQAD)

Syntax

Basic expressions. As L(GQA) plus:

1. Conditional symbols: C. Examples: *usually*, *always*.

Derived expressions. As L(GQA).

Formation rules. As L(GQA) plus:

D0. If F is a closed formula of L(GQA), then F is a formula of L(GQAD).

D1. If F_1 and F_2 are formulae, then $F_1; F_2$ is a formula.

D2. If F_1 and F_2 are formulae, and C is a conditional symbol, then $C(F_1, F_2)$ is a formula.

D3. If F_1 and F_2 are formulae, C is a conditional symbol, and i is an index, then $C(i)(F_1, F_2)$ is a formula.

D4. If F is a formula, then $Hyp(F)$ is a formula.

Interpretation

$I(F,V) \; = \; \{W \in V \mid [\![F]\!]^W\}$

Sequencing (alternative definitions):

1. $I(F_1; F_2, V) \; = \; I(F_1, V) \cap I(F_2, V)$
2. $I(F_1; F_2, V) \; = \; I(F_2, I(F_1, V))$
3. $I(F_1; F_2, V) = \begin{cases} I(F_1, V) \cap I(F_2, V) & \text{if } I(F_1, V) \neq \emptyset \\ I(F_2, V) & \text{if } I(F_1, V) = \emptyset \end{cases}$

Conditionals (alternative definitions):

1. $I(C(F_1, F_2), V) \; = \; \{W \in V \mid C'(I(F_1, V), I(F_1, V) \cap I(F_2, V))\}$
2. $I(C(F_1, F_2), V) \; = \; \{W \in V \mid C'(\sum I(F_1, V), \sum I(F_2, I(F_1, V)))\}$

$I(C(i)(F_1, F_2), V) \; = \; C'(\sum I(F_1, V){\downarrow}i, \sum I(F_2, I(F_1, V)){\downarrow}i)$

$I(Hyp(F), V) = \{W \in V \mid I(F, V)\}$

B.4 English fragment

The sentence and discourse grammars are combined here.

Syntactic categories

1. S (sentence)
2. NP (noun phrase), N′ (nominal), N (noun), PN (proper noun), Pro (pronoun)
3. VP (verb phrase), IV (intransitive verb), TV (transitive verb), DV (ditransitive verb)
4. Rel (relative pronoun)
5. Num (numeral)
6. Conj (conjunction)
7. D (discourse).
8. Adv (adverb of quantification).

Production rules

S1. S:qs $\rightarrow$ NP:q VP:s
S2. NP:$p(i)$ $\rightarrow$ PN:p
S3. NP:$p(i)$ $\rightarrow$ Pro:p
S4. NP:$(dp)(i)$ $\rightarrow$ Det:d N$'$:p
S5. NP:$\lambda s_1.q(s_2 \bigwedge s_1)$ $\rightarrow$ NP:q Rel VP:s_2
S6. VP:p_s $\rightarrow$ IV:p
S7. VP:$\hat{x}[q\ \hat{y}[p(x,y)]]$ $\rightarrow$ TV:p NP:q
S8. VP:$\hat{x}[q_1\ \hat{y}[q_2\ \hat{z}[p(x,y,z)]]]$ $\rightarrow$ DV:p NP:q_1 NP:q_2
S9. N$'$:$s_1 \bigwedge s_2$ $\rightarrow$ N$'$:s_1 Rel VP:s_2
S10. N$'$:p_s $\rightarrow$ N:p
S11. NP:$(\Phi\ n(p))$ $\rightarrow$ Num:n N$'$:p
S12. D:f $\rightarrow$ S:f
S13. D:$f_1; f_2$ $\rightarrow$ D:f_1 S:f_2
S14. D:$c(f_1, f_2)$ $\rightarrow$ Adv:c S:f_1 S:f_2
S15. D:$c(i)(f_1, f_2)$ $\rightarrow$ Adv:c S:f_1 S:f_2
S16. D:cf_1f_2 $\rightarrow$ S:f_1 Conj:c S:f_2
S17. D:$Hyp(Pre(f_1 \bigvee f_2))$ $\rightarrow$ either S:f_1 or S:f_2.
S18. VP:cp_1p_2 $\rightarrow$ VP:p_1 Conj:c VP:p_2
S19. TV:cp_1p_2 $\rightarrow$ TV:p_1 Conj:c TV:p_2
S20. DV:cp_1p_2 $\rightarrow$ DV:p_1 Conj:c DV:p_2
S21. N$'$:cp_1p_2 $\rightarrow$ N$'$:p_1 Conj:c N$'$:p_2
S22. NP:$c(q_1(i_1))(q_2(i_2))$ $\rightarrow$ NP:q_1 Conj:c NP:q_2
S23. S:$\neg(f)$ $\rightarrow$ IINTT S:f
S24. S:$\neg(qs)$ $\rightarrow$ NP:q do not VP:s
S25. VP:$\neg(s)$ $\rightarrow$ do not VP:s

B.5 Additional rules

B.5.1 Translation principles

Variant translation principle (VTP).

If a NP whose determiner is *a* is translated as a quantifier occurring in the restriction or body of a determiner D, then it may be translated using the appropriate variant specified in the lexical entry for D.

Pronoun Variant Translation Principle (PVTP).

A pronoun is translated as $(the\ Of_r(i, un))(j)$, if:

The antecedent translates as a quantifier Q_1 and the pronoun as a quantifier Q_2 such that

EITHER Q_1 combines with a set term S containing Q_2

OR there is a quantifier Q which immediately encloses Q_1 and which combines with a set term containing Q_2.

When the conditions are satisfied, the antecedent and pronoun must agree syntactically.

Quantifier scope rules.

1. A formula of the form
   $Q_1(i)\ \hat{x}[Q_2(j)\ \hat{y}[F]]$
   may be replaced by
   $Q_2(j)\ \hat{y}[Q_1(i)\ \hat{x}[F]]$
   provided there are no free occurrences of x in Q_2 or of y in Q_1.
2. A formula of the form
   $(D_1\ (S_1 \wedge\ \hat{x}[Q(j)\ \hat{y}[F]]))(i)\ S_2$
   may be replaced by
   $Q(j)\ \hat{y}[(D_1(S_1 \wedge\ \hat{x}[F]))(i)\ S_2]$
   provided there are no free occurrences of x in Q or of y in Q_1.

Widening rule.

Given a formula F containing a quantifier Q with index j, F may be replaced by $Q(j)\ \hat{u}[F']$, where u is a variable that does not occur in F and F' is obtained from F by replacing Q with $(the\ Of_r(j, un))$. The rule may only apply if Q contains variables which are free in its new position.

Existence presupposition rule.

Given a formula containing a quantifier Q with index i derived from an antecedent, and a quantifier Q' with input index i derived from an anaphor, apply the widening rule to Q in such a way that the result is $Q\ \hat{u}[S]$, where Q' occurs within S.

B.5.2 Preference rules

Preference rule 1 (indices) Assign anaphor translations the same input and output indices by preference.

Preference rule 2 (variant translations) Apply the variant translation principle by preference.

Preference rule 3 (non-intersective determiners, I) If a noun phrase is translated as a quantifier Q with index i applied to a set term S, i.e. as a formula of the form $Q(i)\ S$, and S contains a quantifier with index j, and there are anaphors with indices i and j which occur in the same sentence as each other, then prefer the intersective translation for the determiner of the noun phrase.

Preference rule 4 (non-intersective determiners, II) If preference rule 3 does not apply, the preferred translation of an upward monotone determiner is the intersective one; of a downward monotone determiner, the non-intersective one.

Preference rule 5 (non-intersective determiners II, revised) If preference rule 3 does not apply, and the antecedent and anaphor occur in sentences of the same monotonicity, then the preferred translation of an upward monotone determiner is the intersective one; of a downward monotone determiner, the non-intersective one. In sentences of opposite monotonicity the reverse applies.

References

Nicholas Asher (1986). Belief in Discourse Representation Theory. *Journal of Philosophical Logic*, 15:127–189.

Nicholas Asher (1987). A Typology for Attitude Verbs and their Anaphoric Properties. *Linguistics and Philosophy*, 10:125–197.

Nicholas Asher (1990). Three Problems for the Semantics of Gerundive Constructions. In Martin Stokhof and Leen Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, pages 1–27, ITLI, Amsterdam.

Damaris M. Ayuso (1989). Discourse Entities in Janus. In *27th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 243–250.

R. Baüerle and U. Egli (1985). *Anapher, Nominalphrase und Eselssätze.* Papier 105 des Sonderlorschungbereichs 99, Universität Konstanz.

R. Bartsch (1987). Frame Representations And Discourse Representations. *Theoretical Linguistics*, 14:65–117.

Jon Barwise and Robin Cooper (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4:159–219.

Jon Barwise and John Perry (1983). *Situations and Attitudes.* MIT Press (Bradford Books), Cambridge, Ma.

Jon Barwise (1987). Noun Phrases, Generalized Quantifiers, and Anaphora. In Peter Gärdenfors (ed.), *Generalized Quantifiers*, pages 1–29, D. Reidel, Dordrecht.

David Beaver (1991). DMG Through the Looking Glass. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 135–154.

Peter Bosch (1983). *Agreement and Anaphora: A Study of the Role of Pronouns in Syntax and Discourse.* Academic Press, London.

David Carter (1987). *Interpreting Anaphors in Natural Language Text.* Ellis Horwood, Chichester.

Gennaro Chierchia and Mats Rooth (1984). Configurational Notions in Discourse Representation Theory. *Proceedings of NELS*, 14:49–63.

Gennaro Chierchia (1991). Anaphora and Dynamic Logic. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 37–78.

Herbert H. Clark and Eve V. Clark (1977). *Psychology of Language: An Introduction to Psycholinguistics.* Harcourt Brace Javanovich, New York.

Robin Cooper and Hans Kamp (1991). Negation in Situation Semantics and Discourse Representation Theory. In Jerry Seligman (ed.), *Partial and Dynamic Semantics II, DYANA Deliverable R2.1.B*, pages 61–80.

Robin Cooper (1979). The Interpretation of Pronouns. In F. Heny and H. S. Schnelle (eds.), *Syntax and Semantics 10: Selections from the Third Groningen Round Table*, pages 61–92, Academic Press, New York.

Robin Cooper (1991). Donkeys and Situations. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 155–175.

Paul Dekker (1991). The Scope of Negation in Discourse. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 79–134.

David R. Dowty, Robert E. Wall, and Stanley Peters (1988). *Introduction to Montague Semantics. Studies in Linguistics and Philosophy 11*, D. Reidel, Dordrecht.

David Dowty (1988). Type Raising, Functional Composition and Non-constituent Coordination. In Richard T. Oehrle, Emmon Bach, and Deidre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, pages 153–197, D. Reidel, Dordrecht.

Gareth Evans (1977). Pronouns, Quantifiers and Relative Clauses. *Canadian Journal of Philosophy*, 7:467–536. Reprinted in G. Evans, *Collected Works*, Clarendon Press, Oxford, 1985.

Gareth Evans (1980). Pronouns. *Linguistic Inquiry*, 11:337–362.

Werner Frey and Uwe Reyle (1983). A Prolog Implementation of Lexical Functional Grammar as a Base for a Natural Language Understanding System. In *1st Conference of the European Chapter of the Association for Computational Linguistics: Proceedings of the Conference*, pages 52–57.

Jean Mark Gawron and Stanley Peters (1990). *Anaphora and Quantification in Situation Semantics. CSLI Lecture Notes 19*, CSLI, Stanford.

Peter Thomas Geach (1962). *Reference and Generality: An Examination of Some Medieval and Modern Theories.* Cornell University Press, London.

Brendan S. Gillon (1987). The Readings of Plural Noun Phrases in English. *Linguistics and Philosophy*, 10:199–219.

Jeroen A. G. Groenendijk and Martin J. B. Stokhof (1987). Dynamic Predicate Logic: Towards a Compositional and Non-representational Discourse Theory. ms, ITLI, Amsterdam.

Jeroen A. G. Groenendijk and Martin J. B. Stokhof (1988). Context and Information in Dynamic Semantics. In B. Elsendoorn and H. Bouma (eds.), *Working Models of Human Perception*, Academic Press, New York.

Jeroen Groenendijk and Martin Stokhof (1991a). Dynamic Montague Grammar. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 1–35.

Jeroen Groenendijk and Martin Stokhof (1991b). Dynamic Predicate Logic. *Linguistics and Philosophy*, 14:39–100.

Barbara J. Grosz (1986). The Representation of Focus in a System for Understanding Dialog. In B. Grosz, K. S. Jones, and B. L. Webber (eds.), *Readings in Natural Language Processing*, pages 353–362, Morgan Kauffman, Los Altos, California.

Franz Guenthner and Hubert Lehmann (1983). Rules for Pronominalization. In *1st Conference of the European Chapter of the Association for Computational Linguistics: Proceedings of the Conference*, pages 144–151.

F. Guenthner (1987). Linguistic Meaning In Discourse Representation Theory. *Synthese*, 73:569–598.

Nicholas J. Haddock (1988). *Incremental Semantics and Interactive Syntactic Processing.* PhD thesis, University of Edinburgh.

Irene R. Heim (1982). *The Semantics of Definite and Indefinite Noun Phrases.* PhD thesis, University of Massachusetts, Amherst.

Irene Heim (1983). File Change Semantics and the Familiarity Theory of Definiteness. In Rainer Bauerle, Christoph Schwarze, and Arnim von Stechow (eds.), *Meaning, Use and Interpretation of Language*, pages 164–189, de Gruyter, Berlin.

Irene Heim (1988). *The Semantics of Definite and Indefinite Noun Phrases.* Garland, New York.

Irene Heim (1990). E-type Pronouns and Donkey Anaphora. *Linguistics and Philosophy*, 13:137–177.

Graeme Hirst (1981). *Anaphora in Natural Language Understanding: A Survey. Lecture Notes in Computer Science 119*, Springer-Verlag, Berlin.

Jerry R. Hobbs (1986). Resolving Pronoun References. In B. Grosz, K. S. Jones, and B. L. Webber (eds.), *Readings in Natural Language Processing*, pages 339–352, Morgan Kauffman, Los Altos, California.

Mark Johnson and Ewan Klein (1986). Discourse, Anaphora and Parsing. In *11th International Conference on Computational Linguistics: Proceedings*, pages 669–675, COLING, IKS, Bonn.

P. N. Johnson-Laird (1983). *Mental Models.* Cambridge University Press, Cambridge.

Nirit Kadmon (1985). The Discourse Representation of Noun Phrases with Numeral Determiners. *Proceedings of NELS*, 15:207–219.

Nirit Kadmon (1990). Uniqueness. *Linguistics and Philosophy*, 13:273–324.

László Kálmán and Zoltán Szabó (1990). D.I.R.T.: An Overview. In Martin Stokhof and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, pages 253–275, ITLI, Amsterdam.

Hans Kamp and Uwe Reyle (1990). From Discourse to Logic. Course notes issued at the Second European Summer School in Language, Logic and Information (Themes in Discourse Representation Theory).

Hans Kamp and Christian Rohrer (1983). Tense in Texts. In Rainer Bauerle, Christoph Schwarze, and Arnim von Stechow (eds.), *Meaning, Use and Interpretation of Language*, pages 250–269, de Gruyter, Berlin.

Hans Kamp (1981). A Theory of Truth and Semantic Representation. In J. A. G. Groenendijk, T. M. V. Janssen, and M. J. B. Stokhof (eds.), *Formal Methods in the Study of Language*, pages 277–322, Mathematish Centrum, Amsterdam.

Hans Kamp (1984). A Theory of Truth and Semantic Representation. In J. A. G. Groenendijk, Theo M. V. Janssen, and Martin Stokhof (eds.), *Truth, Interpretation and Information: Proceedings of the Third Amsterdam Colloquium*, pages 1–41, Foris, Dordrecht.

Hans Kamp (1984/85). Context, Thought and Communication. *Proceedings of the Aristotelian Society*, 85:239–261.

Hans Kamp (1990). Comments on Groenendijk and Stokhof. In Johan van Benthem (ed.), *Partial and Dynamic Semantics I, DYANA Deliverable R2.1.A*, pages 108–131.

Hans Kamp (1991a). On the Representation and Transmission of Information. In E. Klein and F. Veltman (eds.), *Natural Language and Speech*, pages 135–158, Springer-Verlag, Berlin.

Hans Kamp (1991b). Uniqueness Presuppositions and Plural Anaphora in DTT and DRT. In Martin Stokhof, Jeroen Groenendijk, and David Beaver (eds.), *Quantification and Anaphora I, DYANA Deliverable R2.2.A*, pages 177–190.

Ruth M. Kempson and Annabel Cormack (1981). Ambiguity and Quantification. *Linguistics and Philosophy*, 4:259–309.

Ewan Klein (1986). VP Ellipsis in DR Theory. In Jeroen Groenendijk, Dick de Jongh, and Martin Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, pages 161–187, Foris, Dordrecht.

Fred Landman (1987). Groups. Department of Linguistics, University of Massachusetts.

Alex Lascarides and Nicholas Asher (1991). Discourse Relations and Defeasible Knowledge. In *29th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 55–62.

H. Lasnik (1976). Remarks on Coreference. *Linguistic Analysis*, 2:1–22.

Stephen C. Levinson (1983). *Pragmatics.* Cambridge University Press, Cambridge.

D. Lewis (1975). Adverbs of Quantification. In E. Keenan (ed.), *Formal Semantics of Natural Language*, pages 3–15, Cambridge University Press, Cambridge.

Godehard Link (1983). The Logical Analysis of Plurals and Mass Terms: A Lattice-theoretic Approach. In Rainer Bauerle, Christoph Schwarze, and Arnim von Stechow (eds.), *Meaning, Use and Interpretation of Language*, pages 302–323, de Gruyter, Berlin.

Godehard Link (1984a). Hydras: On the Logic of Relative Clause Constructions with Multiple Heads. In Fred Landman and Frank Veltman (eds.), *Varieties of Formal Semantics: Proceedings of the Fourth Amsterdam Colloquium*, pages 245–257, Foris, Dordrecht.

Godehard Link (1984b). Plural. In A. von Stechow and D. Wunderlich (eds.), *Handbook of Semantics*, to appear.

Godehard Link (1987). Generalized Quantifiers and Plurals. In Peter Gärdenfors (ed.), *Generalized Quantifiers*, pages 151–180, D. Reidel, Dordrecht.

Jan Tore Lønning (1987). Collective Readings of Definite and Indefinite Noun Phrases. In Peter Gärdenfors (ed.), *Generalized Quantifiers*, pages 203–235, D. Reidel, Dordrecht.

Jan Tore Lønning (1991). Among Readings: Some Comments on 'Among Collections'. In Jaap van der Does (ed.), *Quantification and Anaphora II, DYANA Deliverable R2.2.B*, pages 37–51.

Douglas B. Moran and Fernando C. N. Pereira (1992). Quantifier Scoping. In Hiyan Alshawi (ed.), *The Core Language Engine*, pages 149–172, MIT Press (Bradford Books), Cambridge, Ma.

Reinhard Muskens (1991). Anaphora and the Logic of Change. In Jan van Eijck (ed.), *Logics in AI*, pages 412–428, Springer-Verlag, Berlin.

Stephen Neale (1990). *Descriptions.* MIT Press, London.

Barbara H. Partee (1984). Nominal and Temporal Anaphora. *Linguistics and Philosophy*, 7:243–286.

Tanya Reinhart (1983). *Anaphora and Semantic Interpretation.* Croom Helm, London.

Uwe Reyle (1985). Grammatical Functions, Discourse Referents and Quantification. In *Proceedings of IJCAI 9*, pages 829–831.

Barry Richards (1984). On Interpreting Pronouns. *Linguistics and Philosophy*, 7:287–324.

Craige Roberts (1987). *Modal Subordination, Anaphora, and Distributivity.* PhD thesis, University of Massachusetts, Amherst.

Craige Roberts (1989). Modal Subordination and Pronominal Anaphora in Discourse. *Linguistics and Philosophy*, 12:683–721.

Mats Rooth (1987). Noun Phrase Interpretation in Montague Grammar, File Change Semantics and Situation Semantics. In Peter Gärdenfors (ed.), *Generalized Quantifiers*, pages 237–268, D. Reidel, Dordrecht.

Bertrand Russell (1905). On Denoting. *Mind*, 14:479–493.

Remko Scha (1984). Distributive, Collective and Cumulative Quantification. In Jeroen Groenendijk, Theo M. V. Janssen, and Martin Stokhof (eds.), *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, pages 131–158, Foris, Dordrecht.

Remko Scha (1991). Afterthoughts on Collections. In Jaap van der Does (ed.), *Quantification and Anaphora II, DYANA Deliverable R2.2.B*, pages 53–58.

Lenhart K. Schubert and Francis Jeffry Pelletier (1989). Generically Speaking, or, Using Discourse Representation Theory to Interpret Generics. In Gennaro Chierchia, Barbara H. Partee, and Raymond Turner (eds.), *Properties, Types and Meaning, II*, pages 193–268, Kluwer Academic Publishers, Dordrecht.

John Searle (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press.

C. Sedogbo (1988). SYLOG: A DRT System in Prolog. In V. Dahl and P. Saint-Dizier (eds.), *Natural Language Understanding and Logic Programming, II*, pages 185–201, North-Holland, Amsterdam.

Peter Sells (1985). *Restrictive and Non-restrictive Modification.* Report CSLI-85-28, CSLI, Stanford.

Stuart M. Shieber (1986). *An Introduction to Unification-based Approaches to Grammar. CSLI Lecture Notes 4*, CSLI, Stanford.

Candace L. Sidner (1986). Focussing in the Comprehension of Definite Anaphora. In B. Grosz, K. S. Jones, and B. L. Webber (eds.), *Readings in Natural Language Processing*, pages 363–394, Morgan Kauffman, Los Altos, California.

Lesley Stirling (1988). *Switch-reference and Logophoricity in Discourse Representation Theory.* PhD thesis, University of Edinburgh.

Peter F. Strawson (1950). On Referring. *Mind*, 59:320–344.

Gregory T. Stump (1984). *The Semantic Variability of Absolute Constructions.* D. Reidel, Dordrecht.

Johan van Benthem (1989). Polyadic Quantifiers. *Linguistics and Philosophy*, 12:437–464.

Kees van Deemter (1990). Forward References in Natural Language. In Martin Stokhof and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, pages 79–94, ITLI, Amsterdam.

Martin van den Berg (1990). A Dynamic Predicate Logic for Plurals. In Martin Stokhof and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, pages 29–51, ITLI, Amsterdam.

Martin van den Berg (1991). Dynamic Generalized Quantifiers. In Jaap van der Does and Jan van Eijck (eds.), *Generalized Quantifier Theory and Applications*, pages 223–244, Dutch Network for Language, Logic and Information.

Jaap van der Does (1991). Among Collections. In Jaap van der Does (ed.), *Quantification and Anaphora II, DYANA Deliverable R2.2.B*, pages 1–35.

Jaap van der Does (1992). *Applied Quantifier Logics.* PhD thesis, University of Amsterdam.

Rob van der Sandt (1990). Anaphora and Accommodation. Course notes issued at the Second European Summer School in Language, Logic and Information, Leuven (Quantification and Anaphora).

Jan van Eijck (1988). Quantification. In A. von Stechow and D. Wunderlich (eds.), *Handbook of Semantics*, to appear.

Henk Verkuyl and Jaap van der Does (1991). The Semantics of Plural Noun Phrases. In Jaap van der Does and Jan van Eijck (eds.), *Generalized Quantifier Theory and Applications*, pages 403–441, Dutch Network for Language, Logic and Information.

Hajime Wada and Nicholas Asher (1986). BUILDRS: An Implementation of DR Theory and LFG. In *11th International Conference on Computational Linguistics: Proceedings*, pages 540–545, COLING, IKS, Bonn.

Thomas Alexander Wasow (1975). Anaphoric Pronouns and Bound Variables. *Language*, 51:368–383.

Bonnie Lynn Webber (1979). *A Formal Approach to Discourse Anaphora.* Garland Publishing Inc., New York.

Bonnie Lynn Webber (1981). Discourse Model Synthesis: Preliminaries to Reference. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag (eds.), *Elements of Discourse Understanding*, pages 283–299, Cambridge University Press, Cambridge.

B. L. Webber (1983). So What Can We Talk About Now. In M. Brady and B. Berwick (eds.), *Computational Models of Discourse*, pages 331–370, MIT Press, Cambridge, Ma.

Dag Westerståhl (1989). Quantifiers in Formal and Natural Languages. In D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, pages 1–131, D. Reidel, Dordrecht.

Henk Zeevat (1986). A Treatment of Belief Sentences in Discourse Representation Theory. In Jeroen Groenendijk, Dick de Jongh, and Martin Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, pages 189–215, Foris, Dordrecht.

Henk Zeevat (1989). A Compositional Approach to Discourse Representation Theory. *Linguistics and Philosophy*, 12:95–131.

Hendrik Willem Zeevat (1991). *Aspects of Discourse Semantics and Unification Grammar.* PhD thesis, University of Amsterdam.

Joost Zwarts (1990). Kinds and Generic terms. In Martin Stokhof and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, pages 685–706, ITLI, Amsterdam.