DTIC
SELECTE
MAR 1 3 1992
S C D

②

Scope Ambiguity and Inference

Massimo Poesio

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

92-06553

92 3 12 009

Scope Ambiguity and Inference

Massimo Poesio

The University of Rochester
Computer Science Department
Rochester, New York   14627

Technical Report 389

July 1991

Abstract

Relational Semantics can be used to give a denotation to the non-disambiguated logical forms used by Natural Language Processing systems, representations in which the quantifiers are left "in situ". Giving a semantics to these logical forms makes it unnecessary for the system to compute all the disambiguated interpretations of a sentence before storing its representation in the knowledge base. Rules of inference can be defined so that the disambiguation process can be formally modeled in a declarative way. 'Weaker' rules of inference can also be specified so that conclusions can be derived from the non-disambiguated representation.

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TR 389		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Scope Ambiguity and Inference	technical report
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Massimo Poesio	N00014-82-K-0193

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Computer Science Dept. University of Rochester, 734 Comp. Studies Bldg. Rochester, NY, 14627, USA	

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209	July 1991
	13. NUMBER OF PAGES
	45

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
Office of Naval Research Information Systems Arlington, VA 22217	unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

natural language processing; semantics; scope ambiguity; relational semantics; DRT; logical forms

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Relational semantics can be used to give a denotation to the non-disambiguated logical forms used by natural language processing systems, representations in which the quantifiers are left "in situ." Giving a semantics to these logical forms makes it unnecessary for the system to compute all the disambiguated interpretations of a sentence before storing its representation in the knowledge base. Rules of inference can be defined so that the disambiguation process can be formally modeled in a declarative way. 'Weaker' rules of inference can also be specified so that conclusions can be derived from the non-disambiguated representation.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

Contents

1 The Problem

In using (1), a speaker could mean that there is some one undergraduate who is dating all male students, or merely that all male students date some undergraduate or other.

(1)   Every male student dates an undergrad.

The conventional view is that each sentence with multiple interpretations is to be seen as ambiguous, that is, each interpretation has to be represented by a distinct formula. The two interpretations of (1) are represented by (2a) and (2b).

(2)
   a. $((\forall x)\,(\mathrm{MS}(x) \supset ((\exists y)\,\mathrm{U}(y) \wedge \mathrm{D}(x, y))))$
   b. $((\exists y)\,(\mathrm{U}(y) \wedge ((\forall x)\,(\mathrm{MS}(x) \supset \mathrm{D}(x, y)))))$

"Traditional"[1] Natural Language Processing (NLP) systems, such as TEAM [Grosz *et al.*, 1987] or the Core Language Engine [Alshawi *et al.*, 1988], built according to this view, analyze (1) more or less as follows: First, the parser computes a *logical form* [Webber, 1978; Schubert and Pelletier, 1982; Allen, 1987; Alshawi and van Eijck, 1989] which is similar to the S-structure representation of (1) before Quantifier Raising [May, 1985]:

(3)   [<every x male-student> dates <a y undergrad>]

All the unambiguous interpretations of (1) are then extracted from (3) by algorithms like that proposed by Hobbs and Shieber [1987]. Finally, the system must choose an interpretation, which is normally done using preference heuristics [Hurum, 1988b].

The disadvantages of this method have not gone unnoticed [Kempson and Cormack, 1981; Hobbs, 1983; Allen, 1991; Hirst, 1990]. In this thesis I will concentrate on two of the problems discussed in the literature. The first problem is that a system like the one just described cannot use information which comes later in the discourse. Yet, that information could save the system considerable work. Suppose, for example, that sentence (1) is followed by sentence (4), and that the system is able to conclude that *her* is anaphoric to *an undergrad* in (1). It could immediately conclude that *an undergrad* in (1) has wide scope.

(4)   I met her yesterday.

The second problem is that the number of interpretations can be very large, and therefore computing them all can be very expensive. This great number of interpretations is caused by at least two factors: First, the number of scopally distinct interpretations grows with the factorial of the number of NP's, with the result that sentence (5) has 5! = 120 interpretations. Yet, people do not seem to entertain 120 possibilities when hearing (5).

(5) In most democratic countries most politicians can fool most
      of the people on almost every issue most of the time.   [Hobbs, 1983]

---

[1]The use of "traditional" here should not be thought of as derogatory. Actually, it is more synonymous with "working".

3

Scope ambiguities can combine with other forms of ambiguity, and this increases the number of disambiguated interpretations even further. By solely considering that in a sentence like (6a) the examiners may be involved to a different degree in the grading, Kempson and Cormack [1981] are able to find at least four interpretations for it.

(6)   a. Two examiners marked six scripts.
      b. Three Frenchmen visited five Russians.

(6a) can be used to mean (i) that the same six scripts were each marked by two examiners, (ii) that two examiners marked six (not necessarily the same) scripts each, (iii) that two examiners marked a group of six scripts between them, and (iv) that two examiners each marked the same set of six scripts. As for (6b), Partee ([1975], quoted by Bunt [1985]) argues that it has eight readings; Bunt, also counting collective and distributive interpretations, is able to find 30 different readings for it! And then one must take into account lexical ambiguity, referential ambiguity, and so forth. It seems unlikely that people generate all these interpretations when processing (5), (6a) or (6b). Even 30 interpretations seems too large a number to be actually considered. And yet, sentences of this type are more common than normally believed, as shown by sentences (7)-(9), taken from a set of 10 computer news articles:

(7)   There also was, however, no change in the long-term belief by many
      people in the capital – some Republicans as well as nearly all
      Democrats – that Quayle was simply unqualified to become president.

(8)   McGee has used the whistle-blowing technique numerous times
      over the past several years.

(9)   The yacht was often used for social and political events
      by several presidents until Carter disposed of it.

and people don't seem to have trouble with these sentences.

Relational Semantics is a semantic system related to Situation Semantics [Barwise and Perry, 1983] and developed to explicitly represent the different ways in which NP's contribute to discourse [Heim, 1982; Barwise, 1987; Rooth, 1987]. One of the basic ideas of relational semantics is that the denotation of a sentence is not a truth value, but a *relation*, that is, a set of pairs. The first element of each pair is a variable assignment which gives to the variables values which satisfy the discourse prior to the sentence. The second element is a variable assignment which satisfies the discourse after the sentence has been added to it [Barwise, 1987; Rooth, 1987].

If we use relational semantics instead of the classical truth-valued semantics, we can assign a denotation to the "unscoped" expressions found in the NLP literature. The denotation of such an expression will be the union of the relations denoted by each of the disambiguated representations derived from that expression. Why is it useful to give a denotation to these expressions? We can exploit this ability in three ways: First, we can define weaker versions of the standard inference rules, so that certain types of conclusions

4

can be derived without disambiguating. Second, we can define semantically justified inference rules so that context information can be used for a complete or partial disambiguation. Third, we may conceive a declarative way of dealing with the phenomenon of *scoping preferences*, the fact that people prefer certain interpretations over others [Lakoff, 1971; Ioup, 1975; Johnson-Laird, 1977; Kempson and Cormack, 1981; Fodor, 1982; Hurum, 1988a; Moran, 1988; Kurtzman and MacDonald, 1991]. *If desired (and if the logic presented in this paper is extended in the appropriate way)* we can represent these preferences as plausible axioms and model their interaction with the other sources of disambiguating information explicitly.

A *NLP system using the representation I will* propose needn't compute all the interpretations of (1) right away. Once such a system has translated (1) into an unscoped logical form, it can immediately store the logical form in the discourse representation. At this point, the system may decide whether to disambiguate or not according to some sort of utility measure. If, later in the discourse, a sentence like (4) is also asserted, and if the system is able to conclude that *her* in the second sentence is anaphoric to *an undergrad* in the first sentence, it will also be able to conclude that *an undergrad* takes scope over *every male student*. It should be clear that, in this way, both of the problems with the previous kind of architecture are solved.

I will first discuss in section 2 three proposals for dealing with scope ambiguity – the idea of using disjunction, the "radical vagueness proposal" of Kempson and Cormack, and Hobbs' solution based on dependency functions. I will explain why these solutions are all incomplete in one way or the other. I will also discuss an alternative to the method of generating *all the interpretations* – the idea of using preference principles to generate a single interpretation and then backtrack in case a contradiction is found. This solution is not explicitly presented anywhere in the literature, but is implicitly adopted by several systems: I will argue that a semantics of the kind presented in this paper is a necessary prerequisite for implementing a system which works in this way.

In order to make the proposal more precise I will need a discourse representation which makes it easy to talk about certain discourse inferences, particularly anaphora resolution. I will therefore use a style of representation based on Discourse Representation Theory (DRT) [Kamp, 1981], with the hope that this kind of representation will be better known than, say, Dynamic Montague Grammar [Groenendijk and Stokhof, 1990] or Episodic Logic [Schubert and Hwang, 1990]. I will review DRT very briefly in section 3.

I will introduce relational semantics in section 4, and show how one can replace Kamp's semantics for DRT with one based on relational semantics. (After this redefinition, the difference between the form of representation I use and the others I mentioned will reduce considerably.)

Section 5 contains the core idea of the proposal: I will first show in section 5.1 that one can use relational semantics to add a construct called *scope forest* to the logic presented in *section 4.2, so that scopally ambiguous sentences can be given a single representation.* I will also introduce a new set of inference rules. Before introducing the logic more formally, I will give examples of derivations using this logic in section 5.4.

In section 7 I will discuss the implications of the theory, and answer some possible *objections. I will* also consider in more detail the issue of preferences.

5

2 Some Solutions

2.1 Disjunction

The easiest way to represent a sentence with multiple interpretations without loss of information is to represent that sentence as a disjunction of its interpretations. Sentence (1), for example, would be represented by the disjunction (10).

$$(10) \quad \begin{array}{l} ((\forall x)\,(\text{MS}(x) \supset ((\exists y)\,\text{U}(y) \wedge \text{D}(x,y)))) \vee \\ ((\exists y)\,(\text{U}(y) \wedge ((\forall x)\,(\text{MS}(x) \supset \text{D}(x,y))))) \end{array}$$

In this manner, it is possible to take advantage of disambiguation information in the context. This solution has several problems, however. One of these, as Kempson and Cormack point out [Kempson and Cormack, 1981], is the *Mapping Problem*: there are a number of reasons for preferring as the semantic representation of a sentence a logical structure as close as possible to its syntactic structure. (10) isn't such a representation.[2] A second, and, I think, decisive, argument against this method is that it requires all the interpretations to be computed, and therefore does not solve the combinatorial explosion problem.

2.2 Vagueness

Kempson and Cormack [1981] contend that the conventional view is misled. Even if (1) or (6a) have different interpretations (as K&C put it, they are *logically* ambiguous), they claim that those sentences are not *linguistically* ambiguous, that is, they have a single semantic representation. In their view, the representation of a sentence with multiple interpretations is the weakest representation entailed by all interpretations. This proposal works fairly well for sentences such as (1), because the two interpretations of that sentence are not in fact distinct: the reading in which a single undergraduate is dating all male students entails the other. The representation initially proposed by Kempson and Cormack for (1) is (11).

$$(11) \quad \exists M\ \forall m_{m \in M}\ \exists U_1\ \exists u_{u \in U}\ \text{D}(m,u)$$

In order to extend this method to represent sentences like (6a), however, something more drastic is called for, since none of the interpretations of (6a) is entailed by each of the other three; the 120 interpretations of (5b) are also all distinct. In order to give a unique semantic representation to sentences like these, Kempson and Cormack must introduce a second version of the theory, in which a much weaker representation is used. The representation of (6a) is (12a), and the representation of (1) in the second version of the theory is shown in (12b).

$$(12) \quad \begin{array}{l} \text{a. } \exists X_2\ \exists S_6\ \exists x_{x \in X_2}\ \exists s_{s \in S_6}\ \text{M}(x,s) \\ \text{b. } \exists M\ \exists U_1\ \exists m_{m \in M}\ \exists u_{u \in U}\ \text{D}(m,u) \end{array}$$

---

[2]Finding the logical representation for (5) is left as an exercise to the reader.

(12b) says that there is a set of male students and a set of undergrads, and that one male student dated one undergrad. These truth conditions are much too weak a representation of (1): a NLP system using (12b) as the representation for (1) would have to pay a high price to avoid computing all the interpretations.

This is not, however, what Kempson and Cormack have in mind. Their idea is that (12b) is not the final representation of (1), but only the 'basis' from which the real interpretations can be generated by means of two operations:

- *uniformising*: when an existential quantifier follows a universal, reverse their order

- *generalising*: turn an existential quantifier into a universal.

But in this case we are left with something not much different from a 'traditional' system - the extraction operations do the job that in a traditional system would be done by an algorithm like Hobbs and Shieber's (assuming that one can justify these operations semantically, which Kempson and Cormack don't) and the 'filters' that Kempson and Cormack use to choose one interpretation over the rest are not much different from preference heuristics.

2.3 Dependency Functions

The proposal advanced by Hobbs in [Hobbs, 1983] falls in a third class of solutions, all based on the idea of representing scope relations as dependency functions. Hobbs' solution is based on a certain set of assumptions. First, Hobbs wants to use a first-order representation, with variables ranging over sets. Second, he represents determiners as relations between two sets - but the sets he has in mind are not, however, the set of sets denoted by the NP and the set denoted by the VP. He instead paraphrases a sentence like *Most men work* as 'there exists a set s which represents a majority of the set of all men, and for each individual y in s, y works'. This paraphrase becomes, in his representation, the formula (13).

(13) $(\exists s) (\text{MOST}(s, \{x|\text{MAN}(x)\}) \wedge (\forall y)(y \in s \supset \text{WORK}(y)))$

Hobbs' third assumption is that sets have *typical elements*. The typical element of a set s is an individual $\tau(s)$ defined by the following axiom:

(14) $(\forall s)P_s(\tau(s)) \equiv (\forall y)(y \in s \supset P(y))$

Where P_s is a predicate which is like P except that it is also true of $\tau(s)$ iff P is true of all the elements of s. Hobbs' representation for (1) is (something like) (15), which can be read as follows: there is a set m which includes all the male students, a set u which contains one undergrad, and the typical element of m dates the typical element of u.

(15) $(\exists m, m_1, u, u_1) (\text{EVERY}(m, m_1) \wedge A(u, u_1) \wedge \text{MALE-STUDENT}_{m_1}(\tau(m_1))$
$\wedge \text{UNDERGRAD}_{u_1}(\tau(u_1)) \wedge \text{DATES}(\tau(m), \tau(u)))$

Finally, scope relations are represented using *dependency functions*. A dependency function f returns, for each male student x, the set of undergrads that x dates:

7

$$f(x) = \{y|\ \text{UNDERGRAD}(y) \land \text{DATES}(x, y)\}.$$

If the inferencing component discovers that there is a different set u for each element of the set m, u can be viewed as referring to the typical element of this set of sets, and the fact $u = \tau(\{f(x)|x \in m\})$ can be added to the knowledge base. There are two problems with this solution: First of all, as Hobbs points out, the representation in (13) can only be used with monotone increasing determiners, like *most* and *every*. For example, if we were to represent *No man works hard* in Hobbs's representation, we would be able to conclude that no man works, which instead doesn't follow because *no* is not monotone increasing. The second problem, common to other dependency function-based solutions, is that only sentences with two quantifiers can be given a scope-neutral representation, and not, for example, sentences with a quantifier and negation, such as *John doesn't have a car*.

2.4 Preferences and Backtracking

It has often been observed in the psycholinguistic and syntactic literature that people prefer certain interpretations over others [Lakoff, 1971; Ioup, 1975; Johnson-Laird, 1977; Kempson and Cormack, 1981; Fodor, 1982]. For example, most people seem to agree that *Every male student* takes wide scope in (1), and the NP *a kid* takes wide scope in (16).

(16) A kid climbed every tree.

These preferred readings have been explained by stipulating the existence of psychologically motivated principles used when parsing their sentences. According to the *Linear Order Principle*, for example, the preferred scope ordering of quantified phrases matches the left-to-right ordering of the phrases in the sentence. This principle goes back to work by Lakoff [Lakoff, 1971], (which actually claimed that sentences like (1) are unambiguous because of this principle!), and would explain why we get the preferred readings in 1 and (16).

The stronger version of Lakoff's claim is clearly untenable, because of sentences like (17a) and (17b):

(17) a. There was a fish on every plate.   [Chierchia and McConnell-Ginet, 1990]
   b. Every student likes the professor of anthropology.

Ioup [1975] proposed however the following revisions to Lakoff's proposal. She replaced the Linear Order Principle with a *Surface Subject* principle, according to which the surface subject tends to take wide scope, especially if it coincides with the deep subject. She also introduced a second principle, known as the *Quantifier Hierarchy Principle*, according to which quantifiers are organized in a hierarchy according to the ease with which they take wide scope. (For example, quantifiers like *each* or *any* tend to take wider scope more often than *some* or *a few*.) The first principle would explain why the preference for the subject to have wide scope is less pronounced in passive sentences like (18); the second principle would explain why *the professor of anthropology* takes wide scope in (17b).

(18) A tree was climbed by every boy.

8

On the basis of these considerations, one might argue for the following kind of architecture: instead of computing all the interpretations, the system has the Left Order Principle "built in", possibly corrected by the Quantifier Hierarchy Principle and generates only one interpretation, the one which agrees with the principle. If inconsistencies are found, the system backtracks, and computes another interpretation. To my knowledge, this architecture has never been explicitly proposed in the literature, but it looks like a potential solution to the problems with the 'traditional' systems mentioned above.

One can however easily find counterexamples to each of the principles proposed in the literature:

(19) a. Every school in this district commemorates an episode
     that occurred some years ago.
  b. The cost of everything from food to cars can be pushed
     artificially high by greedy retailers.
  c. The teacher has more influence than the parents on most children.

(19a) violates the left order principle, (19c) violates both the left order and the quantifier hierarchy principle, and the preferred interpretation of (19b), (for me and a couple of informants) is the one in which *everything* takes scope over *greedy retailers*, which in turns takes scope over *the cost of* .... I doubt that anybody will have problems understanding this sentence, and yet it violates not only the left-to-right ordering principle of Lakoff, but also the heuristic that NP's with the determiner *the* tend to take wide scope.

Kurtzman and MacDonald run experiments to test which of these principles really affect the choice of the interpretation, and how they interact [Kurtzman and MacDonald, 1991]. Their result was that in the case of active sentences there was good evidence in favor of the linear order principle, although this preference is stronger for the "a ... every" order than for the "every ... a" order (in contrast with the Quantifier Hierarchy Principle). They also seemed to find evidence for the *Thematic Hierarchy Principle*, according to which the NP's filling certain thematic roles (and the agent role above all) tend to take wider scope. Kurtzman and MacDonald didn't find any principle of general applicability, however; they also found that even the Linear Order Principle is not used by all speakers.[3]

The problem is that even if we concede that these principles actually exist, they certainly compete with each other and can be overridden by world knowledge. A sentence is often not perceived as ambiguous because it appears in a context which disambiguates it. Consider, for example, the sentence *Every graduate student has to use an office on the 4th floor*. By itself, the sentence could either mean that there is more than one office, or that all graduate students share the same office. In the appropriate context, however, one or the other of the readings becomes preferred:

(20) We have problems with space this year. Every graduate student has to use an office on the 4th floor

---

[3]It's interesting to note that Kurtzman and McDonald's results do seem to be in contrast with the predictions of approaches based on using the weakest possible interpretation, like Kempson and Cormack's, since people do not seem to favor the weakest interpretation, and actually at times have a definite preference for the strong one.

Given these problems, is not clear how the architecture based on preferences and backtracking can be made to work without a clear and declarative way of formalizing the way these preferences work, that is, without a logic which extends the one I will present in this proposal. The interpretation of (19b) generated by such a system, for example, would be completely different from the preferred interpretation. It's unclear how such a system would be able to use the disambiguating information given by the context without such a formalization; and it's even less clear how such a system would decide when to backtrack, and how it would decide which interpretation to try next.

These objections do not diminish the appeal of the idea of using the preferences and backtracking. One of the goals of my future work is to extend the logic I will present in this paper in such a way that this kind of architecture can be formalized, and then compare the two approaches.

3 A Crash Course on DRT

The version of Discourse Representation Theory (DRT) developed by Hans Kamp [Kamp, 1981] was originally meant to provide (i) a general account of the conditional; (ii) an account of the meaning of indefinite descriptions and (iii) an account of pronominal anaphora, especially in "donkey" sentences, i.e., sentences like (21) and (22).

(21)   Every man who owns a donkey beats it.

(22)   If a man owns a donkey, he beats it.

Kamp (and Heim [Heim, 1982]) attempt to "build the discourse structure into the logic", that is, to relate the constraints on anaphora to the meaning of discourse. In the representation defined in [Kamp, 1981], the traditional formulas of first order logic are replaced by *Discourse Representation Structures* (DRS's), which are pairs $\langle M, C \rangle$, where M is a set of *markers* drawn from some set V, and C a set of *conditions*. For example, sentence (23) is represented in DRT by the DRS in (24).

(23)   Pedro owns a donkey.

(24)
$$\boxed{\begin{array}{l} \text{x y} \\ \hline \text{PEDRO(x)} \\ \text{DONKEY(y)} \\ \text{OWNS(x,y)} \end{array}}$$

This DRS contains the two markers x and y, and a set of *atomic* conditions like DONKEY(y). Other, *complex* conditions composed of nested DRS's are used to represent other connectives and for universal quantification (see below.)

Most of the empirical import of DRT comes from the definition of the DRS *construction rules*, and above all those for the interpretation of NP's and of conditionals. Two of these rules were used to build the DRT representation of sentence (23) above.

proper names rule: if α is a proper name, a new marker u (x in the example above) is added to the *universal* DRS (that is, the one not embedded in any other), and a new atomic condition of the form $\alpha(u)$ (PEDRO(x) in the figure) is to the same universal DRS.

indefinite NPs rule: if a α is an indefinite NP (e.g., *a donkey*) then a new marker u (y in the previous example) is added to the *current* DRS, and a new atomic condition of the form $\alpha(u)$ (DONKEY(y) in the example) is added to the same DRS. I will show shortly how this rule has been defined in this way to explain why indefinites like *a donkey* take an existential reading in sentences like (23), but a universal one in conditional sentences like (21).

11

The purpose of Kamp's move from formulas to DRS's becomes clear when an actual discourse is considered. A DRT representation for discourse (25) is obtained by just adding the conditions for the second sentence to the previously shown DRS representing (23), as shown in (26).

(25)   Pedro owns a donkey. He hates it.

(26)

$$
\boxed{
\begin{array}{l}
x \ y \ u \ v \\[4pt]
\text{PEDRO}(x) \\
\text{DONKEY}(y) \\
\text{OWNS}(x,y) \\[4pt]
u = x \\
v = y \\
\text{HATES}(u,v)
\end{array}
}
$$
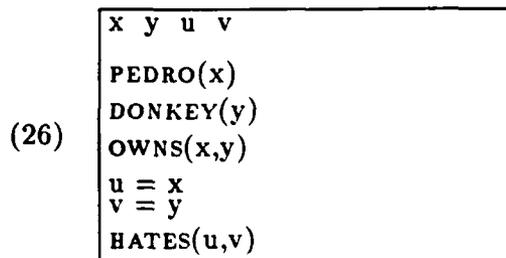
The conditions for *it* are obtained by another DRS construction rule, the *pronoun construction rule*:

pronoun construction rule: If α is a pronoun, introduce a new marker v to the current DRS, choose a *suitable* marker u from the currently *accessible* ones, and add to the current DRS a new condition $v = u$. (A marker u is *accessible* from the DRS K if either u is local to K, or is introduced into a DRS which contains K, as discussed below.)

Every NP's like *every man* in the 'donkey' sentence (21), are handled by the following DRS construction rule:

every construction rule: If *every* α β is a sentence, add to the current DRS a new complex condition of the form $K_1 \Rightarrow K_2$, where K_1 and K_2 are DRS's, adding a new marker u to K_1, adding the conditions for α to K_1, and the conditions for β to K_2.

The results of applying this rule to sentence (21) are shown in (27).

(27)

$$
\boxed{
\begin{array}{l}
\boxed{
\begin{array}{l}
x \ y \\[4pt]
\text{PEDRO}(x) \\
\text{DONKEY}(y) \\
\text{OWNS}(x,y)
\end{array}
}
\ \longrightarrow \
\boxed{
\begin{array}{l}
u \ v \\[4pt]
u = x \\
v = y \\
\text{HATES}(u,v)
\end{array}
}
\end{array}
}
$$
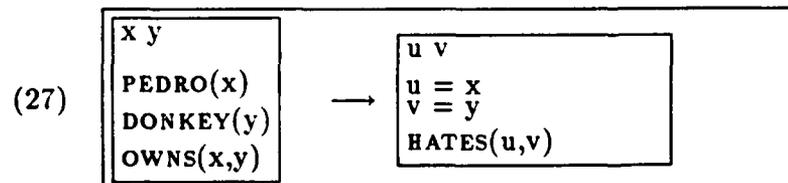
"Accessibility" in DRT is a way of representing, in a "geometrical" fashion, the constraint on anaphora known as *Scope Constraint*, here presented in a formulation due to Heim [1982]:

Scope Constraint (SC): *Do not adjoin an NP any higher than to the lowest S in which it originates.*

This constraint states that no quantified expression can take wider scope than the clause in which it originates. The SC, together with the other usual constraint that no quantified expression can serve as the antecedent of a pronoun outside the scope of that expression, is intended to explain why discourses like (28) are ungrammatical. The scope of *every man* is limited to the first sentence, and therefore it cannot serve as antecedent of *he*.

(28)   *Every man owns a donkey. He hates it.

DRT introduces the "boxes" to make this constraint more apparent. A complete definition of the accessibility conditions, which also takes into account complex DRS's as (27), is as follows: a marker u is *accessible* from the DRS K if either u is local to K, or is introduced in a DRS K' such that $K' \Rightarrow K$, or is introduced in a DRS K' which contains K. This definition predicts that since in the representation of (25) both the markers x and y are in the same DRS to which the conditions for *He hates it* are to be added, they are accessible, and can therefore be used as antecedents for *He* and *it*, respectively. To understand how accessibility and the box notation can be used to explain the ungrammaticality of (28), think of how the DRS in (27) could be extended to represent discourse (28): in analogy to what was done for (25), new markers z and w would be added to the DRS for (21), but there would be no accessible marker to be equated with them, since the markers x and y would be embedded in a DRS not accessible from the outer DRS.

A model for DRT is a structure $\langle U, F \rangle$ with universe U and interpretation function F. F assigns an element of U to each proper name of the language (in this case, English), a subset of U to each of its basic common nouns and basic intransitive verbs, and a set of pairs of elements of U to each basic transitive verb.

Truth is again defined in two stages: First, Kamp introduces the notion of *verification*, which is analogous to the notion of satisfaction in ordinary logic. An *embedding function* is a function $f : V' \mapsto U$, where V' is a subset of V. The embedding function f *verifies* the DRS K, iff dom(f) = the set V_K of markers of K, and for each condition $C_i \in K$, f verifies C_i. Verification for atomic conditions is defined as satisfaction for atomic formulas relative to an assignment. If an atomic condition C is of the form $\alpha(\beta)$, $\alpha = \beta$ or $\alpha(\beta, \gamma)$, that is, of the form $R(\alpha_1, \ldots \alpha_n)$, f verifies C iff $\langle f(\alpha_1) \ldots f(\alpha_n) \rangle \in F(R)$.

The verification conditions for a complex condition of the form $K_1 \Rightarrow K_2$ are defined as follows. Let us first extend the V_K notation used before to embedded DRS's: if K is the 'universal' DRS V_K is the set of markers 'introduced' in K, else V_K is the set of markers introduced in K together with all the markers introduced in the DRS's in K is embedded. An embedding f verifies a condition of the form $K_1 \Rightarrow K_2$ iff for every embedding g such that g verifies K_1 (which means that dom(g) = V_{K_1}, and g verifies all the conditions in K_1) there is an embedding h which verifies K_2. It's easy to verify that this definition gives to the indefinite *a donkey* in (21) the desired universal force.

Informally, a marker is free in a DRS K if it is used in some condition but not 'listed at the top'. A DRS without free markers K is then defined to be true wrt a model M = $\langle U, F \rangle$ iff there is some embedding function f with values in U such that f verifies K. At this point, one can define logical truth and entailment as usual.

4 Relational Semantics

In this section I will review the main ideas of relational semantics. To make the explanation concrete, I will show how one can specify a relational semantics for a DRT representation. An additional goal of this section is to show that DRT can be used as a logic in the conventional sense, that is, to perform inferences. I will call the resulting logic DRT_0.

4.1 Meanings as Relations

Relational Semantics was originally developed to model explicitly the kinds of constraints that NP's impose on discourse. These constraints can be expressed in terms of the requirements on variable assignments. For example, after the sentence *A farmer with a donkey beat it* has been added to a discourse, every variable assignment f which satisfies the new discourse must assign to the variable used to represent the NP *a farmer* a value not used before. After a sentence like *He sat under the table*, instead, every variable assignment which satisfies the discourse must give to the variable used to represent *He* the same value assigned to some other variable.

A comparison with standard first order logic may help. Consider the sentence *A farmer with a donkey beat it*, and assume that the coindexing relations are those represented by the indices in (29).

(29) [A farmer$_1$ with a donkey$_2$ beat it$_2$]

Sentence (29) is represented in first order logic by (30).

(30) $\exists y \, \exists x \, [\text{FARMER}'(x) \land \text{DONKEY}'(y) \land \text{WITH}'(x,y) \land \text{BEAT}'(x,y)]$

Instead of saying that (30) is true with respect to a model, M, and a variable assignment, f, we can say that the meaning of (30) in a model M is the set of assignments with values in M which satisfy it:

(31) $\| \, [\text{A farmer}_1 \text{ with a donkey}_2 \text{ beat it}_2] \, \|^M = \{ \, f \mid f \text{ satisfies} \ldots \}$

Relational Semantics takes this one step further. The crucial idea is that the constraints imposed on a discourse by the NP's can be modeled most effectively by using *partial* variable assignments and by requiring that each variable assignment which satisfies the whole discourse be an extension of a variable assignment which satisfies the portion of discourse prior to the last sentence. Symbolically, a sentence like (29) will cause an extension in the assignment, as shown in (32):

(32) f [A farmer$_1$ with a donkey$_2$ beat it$_2$] f'

This can be represented by requiring the value of (29) to be not a set of assignments as in (31), but a *relation*, that is, a set of pairs of assignments, where the first element of each pair is a variable assignment which satisfies the discourse prior to (29), and the second element is

a variable assignment which satisfies the discourse after (29) has been added [Barwise, 1987; Rooth, 1987]:

$$(33) \quad \{\langle f, f'' \rangle | \; \exists y \exists x [\text{FARMER}'(x) \wedge \text{DONKEY}'(y) \wedge \text{WITH}'(x,y) \wedge \text{BEAT}'(x,y) \\ \wedge f'' = f_x^1 \, _y^2]\}$$

This idea of 'meaning as relations' is also used in Situation Semantics [Barwise and Perry, 1983] and it is becoming increasingly popular because it is very useful to capture certain properties of anaphoric relations [Schubert and Pelletier, 1988; Groenendijk and Stokhof, 1990].

4.2 A Relational Semantics for DRT

The Syntax of DRT$_0$

The set of symbols of DRT$_0$ includes a set of **property symbols** (unary predicates), a set of **relational symbols**, and a set of **markers**: $x_0, \ldots, x_n, \ldots$. (I will sometimes use for simplicity letters without indices like x, y, etc. for the markers.) The set of expressions of DRT$_0$ consists of:

1. **marker introducers** like α^{x_i}, where x_i is a *new* marker, a marker not used for any other marker introducer. (said otherwise, i must be strictly greater than any previously used marker index.)

2. **conditions:**

   (a) **unary conditions** like $P(x_i)$, where x_i is a marker and P is a property symbol.

   (b) **binary conditions** like $R(x_i, x_j)$, where x_i and x_j are markers and R is a relation symbol.

   (c) **coindexing conditions** like $x_i = x_j$, where x_i and x_j are markers.

   (d) **negated DRSs** of the form $\neg K$, where K is a DRS.

   (e) **conditional DRSs** of the form $K_1 \rightarrow K_2$, where K_1 and K_2 are DRSs.

3. **Discourse Representation Structures:** a DRS is an expression containing one or more conditions and zero or more marker introducers, usually written as in (34), where α^{x_0} and α^{x_1} are marker introducers, FARMER(x_0) and DONKEY(x_1) are unary conditions, and OWNS(x_0,x_1) is a binary condition.

(34)
$$\boxed{\begin{array}{l} \alpha^{x_0} \; \alpha^{x_1} \\[4pt] \text{FARMER}(x_0) \\ \text{DONKEY}(x_1) \\ \text{OWNS}(x_0,x_1) \end{array}}$$

In keeping with the standard conventions, I will reserve the symbol K, possibly with subscripts, to indicate DRS's. Subscripted x's like x_0 will always indicate markers. Let a marker x be **free in** K if no marker introducer α^x is in K. A DRT_0 **formula** is a DRS with no free markers. The donkey sentence *Every farmer who owns a donkey beats it* is represented in DRT_0 by (35).

(35)

$$\boxed{\begin{array}{|l|} \hline \alpha^{x_0}\ \alpha^{x_1} \\[4pt] \text{FARMER}(x_0) \\ \text{DONKEY}(x_1) \\ \text{OWNS}\ (x_0, x_1) \\ \hline \end{array}} \rightarrow \begin{array}{|l|} \hline \alpha^{x_2} \\[4pt] x_2 = x_1 \\ \text{BEATS}(x_0, x_2) \\ \hline \end{array}$$

The only significant difference between DRT_0 and 'standard' DRT is the distinction between 'use' and 'introduction' of markers. This distinction makes it easier to enforce the constraint that each marker has to be new, as well as simplifying the definition of the semantics of a DRS, but otherwise has no semantic consequences.

The Semantics

A model M for DRT_0 is a pair $\langle U, F \rangle$: U is a nonempty set, and F an interpretation function. Assignments are called **embedding functions** in DRT; embedding functions are partial functions from markers to objects of the domain. An embedding function over M is a function which associates to the markers values from U. The denotation with respect to M of an expression of DRT_0 is a set of pairs of embeddings over M defined as follows:

1. $\|\alpha^{x_i}\|^M = \{\langle f, g \rangle |\ f \subseteq g,\ x_i \notin \text{DOM}(f) \text{ and } g = f \cup \langle x_i, a \rangle, \text{ for some } a \in U\}$

2. $\|\text{FARMER}(x_i)\|^M = \{\langle f, f \rangle |\ f(x_i) \in F(\text{FARMER})\}$

3. $\|\text{OWNS}(x_i, x_j)\|^M = \{\langle f, f \rangle |\ \langle f(x_i), f(x_j) \rangle \in F(\text{OWNS})\}$

4. $\|x_i = x_j\|^M = \{\langle f, f \rangle |\ f(x_i) = f(x_j)\}$

5. $\left\| \begin{array}{|l|} \hline \alpha^{x_1}, \ldots, \alpha^{x_n} \\[4pt] C_1 \\ \vdots \\ C_m \\ \hline \end{array} \right\|^M = \{\langle f, f \rangle |\ \text{there exist } f_1 \ldots f_n\ \langle f, f_1 \rangle \in \|\alpha^{x_1}\|^M, \ldots$ $\langle f_{n-1}, f_n \rangle \in \|\alpha^{x_n}\|^M \text{ and } \langle f_n, f_n \rangle \in \|C_1\|^M, \ldots \|C_m\|^M\}$

6. $\|\neg K\|^M = \{\langle f, f \rangle | \langle f, f \rangle \notin \|K\|^M\}$

7. $\|K_1 \rightarrow K_2\|^M = \{\langle f, f \rangle |\ \text{for all } f \text{ s.t. for all extensions } g \text{ s.t. } \langle f, g \rangle \in \|K_1\|^M, \text{ there exists } h \text{ s.t. } \langle g, h \rangle \in \|K_2\|^M\}$

It is easy to check that the verification conditions in DRT_0 are analogous to those of standard DRT, and to verify that by requiring that a marker x can only be coindexed with a marker y if the assignment x is defined on both, one also obtains the same accessibility conditions of DRT. Truth can be defined as follows. A formula K is **true** in a model M iff $\|K\|^M \neq \emptyset$. A simple notion of entailment for DRT_0 can be defined as follows: if K_1 and K_2 are formulas, $K_1 \models K_2$ iff for all models M in which K_1 is true, $\|K_1\|^M \subset \|K_2\|^M$.

16

The Inference Rules

I am not aware of any definition of inference rules for DRT in the literature, so I will introduce one that will do for the purposes of this article. A *rule of inference* in DRT_0 is a way of deriving a *conclusion* from a set of *premises*, precisely as in first order logic. That it, the rules of inference of DRT are of the form

$$\frac{K_1,\ldots,K_n}{K}$$

where both the premises $K_1,\ldots,K_n$ and the conclusion K are conditions. The only difference is that these inference rules will be DRS-specific, in the sense that the argument is applicable only when the premises $K_1,\ldots,K_n$ are all conditions of a single *formula* K'; the conclusion will also be added to the same DRS, obtaining a new DRS K''. An inference rule is **acceptable** iff K'' is still a formula. A rule of inference will be **sound** iff it is acceptable, and $K' \models K''$. An example of sound rule of inference for DRT_0 is the following version of *Modus Ponens*:

$$\frac{\boxed{P} \longrightarrow \boxed{Q}, \ P}{Q}$$

5 A Relational Semantics for Unscoped Logical Forms

5.1 A Non-disambiguated Representation

DRT_0 is not a solution to our problems: the two interpretations of sentence (1), in fact, still have to be represented by distinct DRT_0 formulas:

(36)

a.
$$\boxed{\; \boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} \longrightarrow \boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \\ \text{DATES}(x,y) \end{array}} \;}$$

b.
$$\boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \\ \boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} \longrightarrow \boxed{\text{DATES}(x,y)} \end{array}} \; .$$
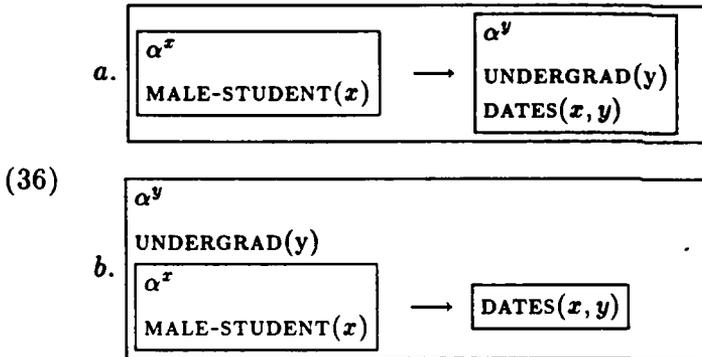
Because of the way the semantics has been defined in section 4.2, however, it is relatively easy to extend DRT_0 with a new construct which can be used to give a unique representation to (1). In this section I will introduce a model of disambiguation in which sentences are represented by **scope forests** whose denotation is the union of the denotations of the scopally disambiguated interpretations, and the number of possible interpretations can be restricted using inference rules which reflect either logical or referential facts.

5.2 Scope Forests

Consider first a slightly modified version of (1).

(37)   Every male student dates most undergrads.

The 'unscoped logical form' representation of (37) proposed in the NLP literature [Schubert and Pelletier, 1982; Allen, 1987; Alshawi and van Eijck, 1989] can be rendered in DRT by something like (38).

(38)
$$\boxed{ < \text{every}_1 \; \boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} > \text{DATES} < \text{most}_2 \; \boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \end{array}} > }$$
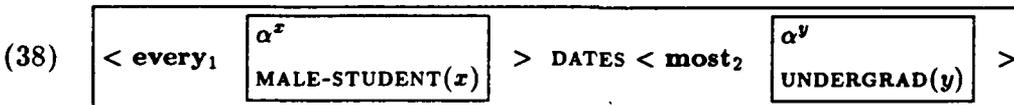
My goal is to give a denotation to a DRS of this type, and I can use for this purpose the relational semantics presented in section 4.2, as follows. There are two quantified NP's in the sentence, *every male student* and *most undergrads*, and two ways of 'ordering' them to get an interpretation. The interpretation of (37) in which the universal takes wide scope (assuming a representation for generalized quantifiers in DRT roughly analogous to the one Kamp suggests in [1988]) is shown in (39).

(39)
$$\boxed{ \boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} \overset{\text{every}_1}{\longrightarrow} \boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \end{array}} \overset{\text{most}_2}{\longrightarrow} \boxed{\text{DATES}(x,y)} }$$
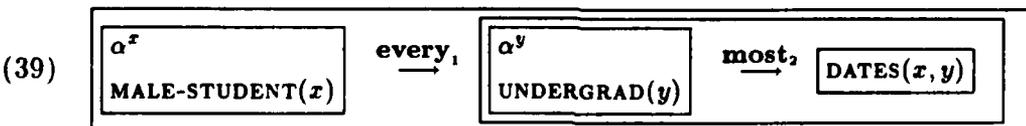
If we think of the restrictions of the quantifiers and of the DRS which represents the scope of *most undergrads* in (39) as nodes of a graph, we can see each way of ordering the quantifiers as a path which starts from the antecedent of the quantifier with wider scope and ends with the consequent of the quantifier with narrower scope. I will therefore call each such way of ordering the NP's a **path**. In the semantics of section 4.2, each path denotes a set of pairs. The denotation of the logical form (38) can therefore be defined as the union of these sets. I will call DRS's like (38) **scope forests**[4]. If the denotation of a scope forest *sf* is the union of the denotations of the paths in a set *ps*, I will say that the paths in *ps* are **associated** with *sf*. The translation rules in the grammar are such that the interpretation of a sentence like (37) is a scope forest[5]. I will call the resulting logic DRT_1.

Ordering Constraints

As new facts about the relative scope of the NP's in a sentence *s* are discovered, the number of paths (that is, interpretations) associated with the scope forest *sf* representing *s* gets smaller. The inference rules for scope disambiguation presented below model this process. These rules use logical truths and facts about reference to derive from a scope forest *sf* a new scope forest *sf'* which has more **ordering constraints**. An ordering constraint of a scope forest *sf* is a label of the form $i < j$, where i and j are indices of operators in *sf*. Only paths in which the operator with index i precedes the operator with index j are associated to a scope forest *sf* labeled with the ordering constraint $i < j$. For example, the scope forest equivalent to the subset of interpretations of (37) in which *most undergrads* takes scope over *every male student* is represented by the DRS in (40).

(40) φ | < $every_1$ | α^x MALE-STUDENT(x) | > DATES < $most_2$ | α^y UNDERGRAD(y) | > $\quad \{2 < 1\}$

Note that only one path is associated to the scope forest (40), which means that the inference rules for disambiguation defined below would apply.

Negation and Indefinites

Before introducing the rules of inference I need to fill in a few details. The first question is how to represent the ambiguities of scope originated by operators like negation. The answer is that the tools introduced so far are sufficient to represent scope ambiguities originated by negation, provided that we also index the negation operator. The representation of the sentence *John doesn't have a car*, for example, will be the scope forest (41). This method can also be used for modal operators.

(41) φ | j $\neg_1$ HAVE < a_2 | α^x CAR(x) | > $\quad \emptyset$

---

[4] This name has historical reasons. 'Scope maze' or something similar would probably be more appropriate

[5] I assume that the logical form is generated as proposed by Schubert and Pelletier [1982].

The second question is how to make the 'path' idea work with sentences like (1), since the representation for indefinites in DRT_0 does not consist of a restriction and a scope. My answer is that it is possible to represent indefinite NP's with structures similar to those used for quantifiers without changing the properties that indefinites have in standard DRT. It is possible, for example, to represent the disambiguated reading of (1) in which *every student* scopes over *an undergrad* as in (42).

(42)

$$\boxed{\;\boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} \quad \overset{\textbf{every}_1}{\longrightarrow} \quad \boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \end{array}} \quad \overset{\textbf{a}_2}{\longrightarrow} \quad \boxed{\text{DATES}(x,y)}\;}$$
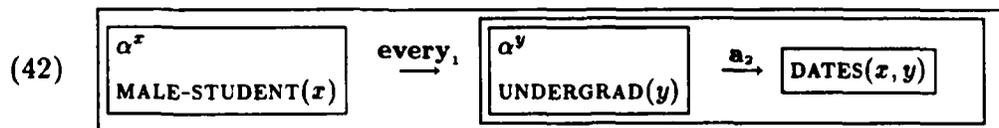
Two properties of indefinites have to be preserved. The first property is that indefinites, unlike quantifiers, are not subject to the scope constraint, as shown by the contrast in acceptability between (43a) and (43b):

(43)   a. A dog_i came in. It_i sat under the table
        b. Every dog_i came in. ??It_i sat under the table

It has been shown however that we can model this distinction semantically, and still represent indefinites with structures like (42) [Rooth, 1987; Schubert and Pelletier, 1988]. We can do this by separating the class of **referential DRS's**[6] used to represent determiners like *a* and *the*, and the pronouns, from the class of **quantified DRS's**, used to represent determiners like **most** and **every**. Both classes of DRS's will have a restriction and a scope, but they will have different semantic properties: in particular, indefinite NP's will have the same accessibility properties that they have in DRT_0.

A second reason for not treating indefinites as quantifiers is that unselective operators like the universal and the conditional seem able to bind indefinites, but not generalized quantifiers. Again, this does not prevent using a representation like (42) in which indefinites have a restriction and a scope. If desired[7], one can achieve the same semantic effects of DRT by giving to generalized quantifiers the capability of imposing constraints on the set of verifying embeddings, as shown in more detail in section 6.

In a word, representing indefinite NP's as in (42) doesn't imply that they get different anaphoric properties, nor that generalized quantifiers have different properties in DRT_1 than they have in DRT_0, unless this change is otherwise motivated. (If desired, we could even define 'simplification rules' for transforming the structures associated to referential NP's into the representation more traditionally associated to indefinites in DRT.) We can therefore use the scope forest notation for (1) as well. The scope forest into which (1) is translated will be as follows:

(44)

$$\left[\; < \textbf{every}_1 \;\boxed{\begin{array}{l} \alpha^x \\ \hline \text{MALE-STUDENT}(x) \end{array}} \quad > \text{DATES} \; (\; \textbf{a}_2 \;\boxed{\begin{array}{l} \alpha^y \\ \hline \text{UNDERGRAD}(y) \end{array}} \quad) \;\right]$$

---
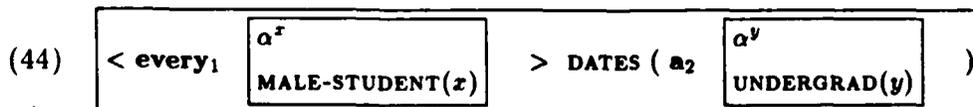
[6] Bad as this name sounds, the only alternative that came to my mind was 'article DRS's', which is even more misleading, because referential DRS's are also used to represent pronouns.

[7] That is, leaving aside the well-known objections raised against the unselective quantification account: the *proportion problem*, presented in more detail in section 6, puts under discussion the claim for generalized quantifiers, while Schubert and Pelletier [Schubert and Pelletier, 1988] present counterexamples to the claim for conditionals.

A remark on the notation: in order to distinguish between referential NP's and *quantifiers in* the scope forest representation, I will use parentheses instead of angle brackets for referential NP's; brackets will be used when any determiner is possible.

Plural Anaphora to Quantifiers

While intersentential singular anaphora to every-NP's and indefinites in the scope of a quantifier is subject to a number of restrictions[8], intersentential plural anaphora is generally possible, as shown by the contrast between (43b) and (45), as well as by the contrast between (46a) and (46b).

(45)   Every dog$_i$ came in. They$_i$ sat under the table.

(46)   a. Every person with a dog$_i$ came in. ??It$_i$ was put under the table.
       b. Every person with a dog$_i$ came in. They$_i$ were put under the table.

Knowing that *they* and *a dog* are anaphorically related in (46b) is useful to disambiguate. I will therefore introduce a formalization of the facts about plural anaphora to singular NP's that, without giving an explanation of the phenomenon, will make it possible to use these facts in the inference rules for disambiguation. I'll borrow the necessary notation from Link's **LP** logic [Link, 1987]. In a model for **LP**, the universe of discourse is not a set, but a complete semilattice $\langle E, \bigvee \rangle$ which contains all the 'sums' of the (atomic) individuals of a set $A \subset E$. An embedding defined over such model can assign to a marker either an atomic individual in A, or an element of $E - A$. I can then introduce the logical predicates ATOM(x), true iff the value associated to x in Link's model is an element of A, and GROUP(x), true iff that value is in $E - A$.[9] An important property of ATOM(x) and GROUP(x) is summarized by the following lemma:

Lemma 5.1 *For every marker x it is either the case that* ATOM(x) *or that* GROUP(x), *but not both.*

The semantics of quantified DRS's like $K_1 \stackrel{\text{every}_i}{\rightarrow} K_2$ will be defined as the result of an operation called **distancing** applied to the set $\{\langle f, g \rangle\}$ of pairs of embeddings such that f verifies the truth conditions of $K_1 \stackrel{\text{every}_i}{\rightarrow} K_2$ and g is one of the embeddings which extend f by giving values to the markers in $K_1 \stackrel{\text{every}_i}{\rightarrow} K_2$ which verify both K_1 and K_2. Embeddings are ways of encoding situations, and distancing can be understood in terms of situations, as a 'change in perspective': in the situations encoded by the embeddings produced from distancing we do not perceive any more the individual events and single objects, but only the situation in its totality and the sets of objects involved. After distancing, only the **projections** of the NP's, that is, the sums of objects playing certain roles in the global situation, are available for discourse anaphora. This is formalized in terms of embeddings

---

[8] Roberts discusses some cases in which it is possible [Roberts, 1987]; see also [Poesio and Zucchi, 1991].

[9] Link's model has already been used by Kamp to represent conjoined and plural NP's in DRT [Kamp, 1988]. Both ATOM and GROUP are mine.

by having distancing return, when applied to a set $\{\langle f, g \rangle\}$ of pairs of embeddings, a new set $\{\langle f, h \rangle\}$ such that h gives to all markers introduced in $K_1 \overset{\text{every}_i}{\rightarrow} K_2$ a value that is the sum of the values given to them by all the extensions of f in the input set. For example, the denotation of the quantified DRS in (42) will be a set $\{\langle f, h \rangle\}$ such that for every pair $\langle f, h \rangle$, $h(x)$ is the sum of all male students which date an undergrad, and $h(y)$ is the sum of all undergrads which are dated by a male student. $h(x)$ is the projection of *every male student*, and $h(y)$ is the projection of *an undergrad*.

According to this account, the contrast between (43a) and (43b), as well as the acceptability of (45), are due to the fact that the projection of *a dog* in (43a) is a unique individual, and therefore available for individual anaphora, while the projection of *every dog* in (43b) and in (45) is the set of all dogs[10].

5.3 Reasoning With Scope Forests

What can we do, then, with a logic with scope forests? First of all, we can do everything that we can do with first order logic, since the normal inference rules (like modus ponens and resolution) are still sound. Second, we can infer certain consequences without disambiguating. Third, we can use the information in the discourse to disambiguate.

### Reasoning without Disambiguating

How can we do inference without disambiguating? In order to do that, we need to define inference rules analogous to first order logic's Universal Instantiation (UI) and Existential Generalization (EG). It's easy to see how such rules can be defined (and semantically justified) in the framework I have been proposing. I will give as an example the scope forest version of Universal Instantiation; 'weak' versions of Existential Generalization and Existential Instantiation can be defined in the same way.

---

[10]This account provides a justification for the operation of *summation* introduced in [Kamp, 1988], but of course doesn't solve the well-known problems raised by (47a) and (47b).

(47) a. Each student walked to the stage. He shook hands with the dean and left.   (Partee)
     b. Each Italian loves his car. He rides it every Sunday.

In [Poesio and Zucchi, 1991] we propose that distancing is blocked if the discourse has a certain structure - for example, the sentence which contains the anaphoric reference is the continuation of an episode along a known 'script'.

WUI (Weak Universal Instantiation) : from the sentences *Every male student dates an undergrad* and *John is a male student* conclude *John dates an undergrad*.

$$\begin{array}{l} \varphi \quad \left< \textbf{every}_k \begin{array}{|c|} \hline \alpha^x \\ \hline P(x) \\ \hline \end{array} \right> \quad R \quad (\textbf{a}_l \begin{array}{|c|} \hline \alpha^y \\ \hline Q(y) \\ \hline \end{array}) \qquad OC \\[4pt] P(b) \end{array}$$

$$\begin{array}{|ccc|} \hline \begin{array}{|c|} \hline \alpha^y \\ \hline Q(y) \\ \hline \end{array} & \xrightarrow{\;\textbf{a}_l\;} & \boxed{R(b,y)} \\ \hline \end{array}$$
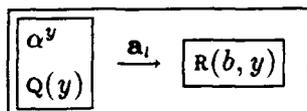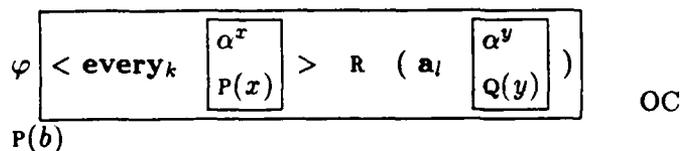
In this rule, as in the rules which I will present in the next section, OC is a set of ordering constraints, and the format is that specified in section 4.2. Similar rules hold when a is replaced by *many*, *most*, etc.

Inference Rules for Scope Disambiguation

We can derive information about the intended scope relations comes from a variety of sources. Three kinds of sources seem especially important:

1. Logical facts, like the fact that the sentence *A male student is dating an undergrad* has only one intepretation.

2. Anaphoric facts: If sentence (1), *Every male student dates an undergrad*, is followed by *They meet them at parties*, and we may conclude that either *them* or *they* is anaphoric to *an undergrad* in (1), that is, we may conclude that the projection of *an undergrad* in (1) is not a single person, but a group of people, then we may also conclude that *every male graduate* scopes over *an undergrad*.

3. World knowledge. For example, one may use facts about the social rules of dating to infer that the most likely interpretation of (1) is the one in which *every student* takes scope over *an undergrad*. As this very example shows, however, most of this information cannot be taken as conclusive, and therefore rules of this type are only appropriate with a logic which allows for revisions.

I will present an example of scope forest reduction rule based on 'logical' facts and two examples of rules based on 'anaphoric' facts. (There is no pretense of completeness: the only reason why I give these specific rules is that they will get the examples in section 5.4 through.) I'll then present the rule to for deriving a disambiguated DRS from a scope forest.

ROR (Referential Over Referential) : This rule reflects the logical fact that referential NP's do not create scope ambiguities: in *A man saw a dog*, for example, the relative scope of *A man* and *a dog* does not matter.

$$\varphi \left(d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; \right) \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC$$

$$\varphi \left(d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; \right) \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC \cup \{i < j\}$$

RAOQ (Referential Atom Over Quantifier) : This rule allows the reduction of the scope forest associated with sentences like *Every male student dates an undergrad* once it has been concluded that *an undergrad* refers to a single individual, that is, the projection of *an undergrad* is an atom in Link's sense. It is worth remembering that because of distancing, y has different values 'inside' and 'outside' of a quantified DRS (therefore, of a scope forest): 'outside' it denotes the *projection* of the referential NP.

$$\varphi \left< d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; > \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC$$
ATOM(y)

$$\varphi \left< d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; > \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC \cup \{j < i\}$$

QORG (Quantifier Over Referential Group) : This rule enables us to conclude, from the fact that the projection of an indefinite NP is a group and the indefinite NP is in a scope forest with a quantifier, that the quantifier takes wide scope. (Consider for example the case when (1) is followed by *They meet them at parties.*)

$$\varphi \left< d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; > \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC$$
GROUP(y)

$$\varphi \left< d_i \; \boxed{\begin{array}{c} \alpha^x \\ P(x) \end{array}} \; > \; R \; \left(d'_j \; \boxed{\begin{array}{c} \alpha^y \\ Q(y) \end{array}} \; \right) \qquad OC \cup \{i < j\}$$

Last but not least, we need to be able to derive a disambiguated DRS from a scope forest associated to a single path. The simplest way for doing this is to introduce a rule of inference

24

whose premise is a scope forest associated to a single path, as follows:

SFE (Scope Forest Elimination) : a scope forest which associated to a single path can be replaced by the corresponding interpretation.

$$\varphi \; \boxed{ < \mathbf{d}_i \; \boxed{\begin{array}{c} \alpha^x \\ \mathrm{P}(x) \end{array}} \; > \quad \mathrm{R} \quad < \mathbf{d}'_j \; \boxed{\begin{array}{c} \alpha^y \\ \mathrm{Q}(y) \end{array}} \; > } \qquad \{i < j\}$$

$$\boxed{\begin{array}{c} \alpha^x \\ \mathrm{P}(x) \end{array}} \; \xrightarrow{\mathbf{d}_i} \; \boxed{\boxed{\begin{array}{c} \alpha^y \\ \mathrm{Q}(y) \end{array}} \; \xrightarrow{\mathbf{d}'_j} \; \boxed{\mathrm{R}(i,j)}}$$

With this way of writing the rules it will be simple to show how the derivations work in section 5.4, but, of course, one would then need one such rule for each permutation of the indices - 120 rules for a sentence with 5 quantifiers, for example. Plus, one would require one such rule for every number of arguments. In a word, it would seem that the combinatorial explosion that I was throwing out of the door is coming back through the window.

It is not so in practice, however. First of all, the particular notation for scope forests I have been using has been chosen to preserve the similarity with the logical forms proposed in the NLP literature as much as possible. With this notation, however, one can only write rules which apply to scope forests with a fixed number of arguments, and these rules are normally asymmetrical, in the sense that the argument position is significant. It should be clear however how rules for n-arguments scope forests could be written, as well as rules in which the argument position is reversed. There are notations in which more general rules can be written, but I preferred not to use them, since they are pretty opaque.

As for the potentially more dangerous problem of requiring $n!$ rules to disambiguate a scope forest with n arguments, a moment's thought will reveal that we don't really need those many rules: all that is really needed is one rule which is triggered when the ordering constraints define a total order (that is, when there is a path of length $n-1$ from an index to another index), and that 'extracts' the operators one after the other from the scope forest. While it is difficult to present such a rule in a simple way, it is not difficult to develop an inference procedure which does this without really going through $n!$ rules. I have a very simple minded algorithm which can add a new constraint and discover if the constraints define a total order in $O(n^2)$. Once it has been determined that the constraints define a total order, the operators can be 'extracted' from the scope forest one after the other in linear time using a procedure similar to the one that will be used in section 6 to define the set of paths associated to a scope forest.

5.4 Reasoning with Scope Forests: Examples

In this section I will show by example what one can do with DRT$_1$. The two examples will show how one can formally derive an unambiguous interpretation from a scope forest using the inference rules presented in the previous section. In this way, I will be able

to take advantage of the examples to show how scope disambiguation can interact with other discourse disambiguation processes, namely, reference disambiguation. I will not give explicit examples of application of WUI; however, the derivations presented below should be explicit enough that the reader should not have problems in reconstructing the derivation of, e.g., *John dates an undergrad*, from the scope forest representation of *Every male student dates an undergrad* and of *John is a male student.* I will first present a simple model of reference disambiguation, then show the derivations.

An Elementary Model of Reference Disambiguation

The set of possible anaphoric antecedents of a pronoun x (its **anchoring set**) initially includes all the markers accessible to x according to a definition of accessibility which is essentially that of DRT, modulo the accessibility of plural markers for quantifiers. The initial anchoring set does not include those markers ruled out by binding constraints [Reinhart, 1983], since the parser introduces a disjointness condition $x \neq y$ for each such marker y.

I will need a logical predicate for talking about accessibility in the object language. The relation between two markers $x_i \prec x_j$ holds whenever x_i is accessible from x_j, that is, whenever x_i is introduced 'before' x_j. Semantically, $x_i \prec x_j$ will be defined to hold whenever $i < j$ (remember that each new marker has a greater index than any of the markers introduced before, and that no two markers are allowed to have the same index) and the embedding gives values to both x_i and x_j.

I will also make the simplifying assumption that if the marker x is the representation of a pronoun, then x is coindexed with one of the accessible markers. This assumption is encoded by the following axiom[11]:

Axiom 5.1 *Let $x_1 \dots x_n$ be all the markers for which $x_j \prec x_{n+1}$ is true, and let x_{n+1} be introduced by a pronoun. Then $x_{n+1} = x_1 \vee \dots \vee x_{n+1} = x_n$ is also true.*

The first reference disambiguation rule adds new disjointness conditions, thus eliminating elements from the anchoring set. Let the $*$ operator be defined as follows:

$$(*OP) \quad P^*(x) \equiv_{def} (\text{ATOM}(x) \wedge P(x)) \vee \left(\text{GROUP}(x) \wedge \boxed{\begin{matrix} \alpha^y \\ y \in x \end{matrix}} \xrightarrow{\text{every}_i} \boxed{P(y)} \right)$$

The ASTR rule says that if two markers x and y are not of the same type, that is, a predicate P is true of the marker x but not of y, then y and x are disjoint.

ASTR (Type Reduction) :

$$\frac{P^*(x), \ \neg P^*(y)}{x \neq y}$$

---

[11]This axiom is too strong in general, but will do for the moment. Consider the sentence *He came in*, and suppose that John and Bill are the only available referents. Using the disjunction method, we would obtain as a representation of the sentence that it is either the case that John came in or that Bill came in. Imagine, however, that neither John nor Bill came in. The theory would then predict that *He came in* is false, which doesn't seem the right prediction: most people would conclude instead that the referent of *he* is neither John nor Bill. This is, I think, yet another argument against using identifying ambiguity with disjunction.

When the anchoring set of x consists of a single element z the **parameter anchoring** (PA) rule applies. Parameter anchoring lets us infer a coindexing relation $x = y$ between a marker x and a marker y whenever every other marker z accessible from x has been found to be distinct from x:

PA (Parameter Anchoring) :

$$y \prec x$$

$$
\boxed{\begin{array}{l} \alpha^z \\ z \prec x \\ y \neq z \end{array}} \xrightarrow[\longrightarrow]{\text{every}} \boxed{x \neq z}
$$

$$\overline{}$$
$$x = y$$

Disambiguation By Deduction

In this section I will show how each disambiguated interpretation of sentence (1) can be derived from the scope forest representation, given the appropriate context. The following theorem (a simple corollary of a lemma presented in section 5.1) will be used in the derivations:

Theorem 5.1 *Given any two markers x and z, if* ATOM(z) *is true and $x = z$ is true, then* ATOM(x) *is also true.*

Let us consider again sentence (1), repeated here for convenience.

(1) Every male student dates an undergrad.

Let us now suppose that sentence (1) is followed by sentence (49).

(49)   I met her.

In Fig. 1 I show how one can deduce a wide scope reading for *an undergrad* in (1) using the fact that it is coindexed with *her* in (49). The first five lines of the derivation are premises, the result of the translation of sentences (1) and (49). Line 6 is obtained from line 3 using the definition of WOMAN*. (Either z is an atom or a group; either way, line 6 is valid.) Line 7 is derived from line 1 using the same reasoning to conclude MALE-STUDENT*(x) and then using world knowledge. At this point we can apply ASTR to derive line 8 from lines 4,6 and 7. Since y is the only marker left in the anchoring set of z, we can apply PA to derive line 9. Line 10 can be derived using the fact that (49) describes a stage level sentence (and some facts about singular anaphora to quantified NP's presented in [Poesio and Zucchi, 1991]). At this point, we can use theorem 5.1 to conclude that the projection of *an undergrad* is also an atom, and therefore we can use RAOQ to derive line 12, from which we can derive line 13 using SFE, if desired.

27

```
1. [<every 1 [x male-student(x)]> dates (a 2 [y undergrad(y)])], 0
                    (translation of 'Every male student dates an undergrad')

2. x <> y              (translation of 'Every male student dates an undergrad',
                        possibly incorporating some form of indefiniteness effect)

3. [speaker met (her 3 [z woman(z)])], 0
                    (translation of 'I met her')

4. x < z

5. y < z

6. woman*(z)           (definition of woman*)

7. ~woman*(x)          (world knowledge)

8. z <> x              (4, 6, 7, ASTR)

9. z = y               (5, 2, 4, 8, PA)

10.atom(z)             (world knowledge)

11.atom(y)             (9, 10, Theorem 5.1)

12.[<every 1 [x male-student(x)]> dates (a 2 [y undergrad(y)])], {2 < 1}
                    (1, 11, RAOQ)

13.[y undergrad(y)] -a-> [[x male-student(x)] -every-> [x dates y]]
                    (12, SFE)
```

Figure 1: Wide scope for *an undergrad*

Next, an example of disambiguation in which *every male student* takes wide scope.
Suppose that (1) is followed by the following sentence:

(50)   They meet them at parties.

The derivation of a wide scope reading for *every male student* is shown in Fig. 2.

```
1. and 2.           as before.

3. [<they 3 [z group(z)]> meet-at-parties <they 4 [w group(w)]>], 0
                    (translation of 'They meet them at parties.')

4. and 5.:          as before

6. z <> w           (translation of 'They meet them at parties.')

7. w = x or w = y   (axiom 5.1)

8. z = x or z = y   (axiom 5.1)

9. y = w or y = z   (6, 7, 8)

10.group(z)         (part of the meaning of 'they')

11.group(w)         (part of the meaning of 'they'

12.group(y)         (9, 10, 11)

13.[<every 1 [x male-student(x)]> dates (a 2 [y undergrad(y)])], {1 < 2}
                    (1, 12, QORG)

14.[x male-student(x)] -every-> [[y undergrad(y)] -a-> [x dates y]]
                    (13, SFE)
```

Figure 2: Narrow scope for *an undergrad*

6 A Formal Presentation of DRT₁

6.1 The Syntax

The main syntactic differences between DRT_0 and DRT_1 are the following:

1. Universally quantified sentences and conditionals are now represented as two different classes of expressions: the class of **quantifier DRSs**, and the class of **connective DRSs**.

2. A new class of complex conditions is introduced, the class of **referential DRSs**, to represent indefinites, definites and pronouns. These kinds of NP's are therefore syntactically separated from proper names.

3. Another new kind of complex condition is introduced, the **scope forest**.

4. Operators like quantifiers and negation are given **indices**.

The set of symbols of DRT_1 includes, in addition to the set of symbols of DRT_0, a set of indices $0 \ldots i \ldots$; a set **QDet** = {**every, most** } of **quantifier operators**; and a set **RDet** = {**a, the, he, it, she, they**} of **referential operators**. The set of expressions of DRT_1 consists of marker introducers (defined as before), conditions, and DRSs. In addition to the unary, binary and coindexing conditions of DRT_0, DRT_1 includes the following types of conditions:

1. **disjointness conditions** of the form $x \neq y$, where x and y are markers.

2. **accessibility conditions** of the form $x \prec y$, where x and y are markers.

3. the **structural conditions** ATOM(x) and GROUP(x), where x is a marker.

4. **negated DRSs**, that are expressions of form $\neg_i K$, where K is a DRS and i an index;

5. **connective DRSs**, like $K_1 \overset{\text{if-then}_i}{\rightarrow} K_2$ (used to represent the conditional), where K_1 and K_2 are DRSs;

6. **quantified DRSs**, which are expressions of the form $K_1 \overset{d_i}{\rightarrow} K_2$, K_1 is of the form

$$\boxed{\begin{array}{l} \alpha^x \\ P(x) \end{array}}$$

   $d \in$ **QDet**, and K_2 is a DRS; K_1 will be called **restriction**, K_2 **scope**, and x the **main marker**;

7. **referential DRSs**, again of the form $K_1 \overset{d_i}{\rightarrow} K_2$, where $d \in$ **RDet**, and K_1 and K_2 are DRSs; again, K_1 will be called restriction and K_2 will be called scope;

8. **scope forests**, which are expressions of one of the following types:

30

$$\varphi \left[< \mathbf{d}_i \; \boxed{\begin{array}{c} \alpha^x \\ \hline P(x) \end{array}} > \; \mathbf{R} \; < \mathbf{d}'_j \; \boxed{\begin{array}{c} \alpha^y \\ \hline Q(y) \end{array}} > \right] \quad OC$$

$$\varphi \left[< \mathbf{d}_i \; \boxed{\begin{array}{c} \alpha^x \\ \hline P(x) \end{array}} > \; \mathbf{R} \right] \quad OC$$
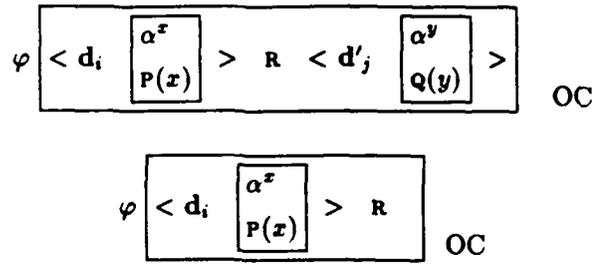
where **d** and **d'** are in **QDet** $\cup$ **RDet**; R is either a relation or a *negated relation* of the form $\neg_i R'$ where R' is a relation; and *OC* is a set $\{ i \prec j, \ldots k \prec l \}$ of **ordering constraints** among the indices of the operators in the scope forest.

A **formula** of DRT$_1$ is a DRS with no free markers, and in which no two operators are given the same index. The donkey sentence *Every farmer who owns a donkey beats it* is represented as in (51) (which the reader should compare to the DRT$_0$ representation (35), section 4.2).

(51)

$$\boxed{\begin{array}{l} \boxed{\begin{array}{l} \alpha^x \\ \hline \text{FARMER}(x) \\ \hline \boxed{\begin{array}{l} \alpha^y \\ \hline \text{DONKEY}(y) \end{array}} \xrightarrow{\mathbf{a_2}} \boxed{\text{OWNS}(x,y)} \end{array}} \quad \overset{\mathbf{every_1}}{\longrightarrow} \; \boxed{\boxed{\begin{array}{l} \alpha^z \\ \hline z = y \end{array}} \overset{\mathbf{it_3}}{\longrightarrow} \boxed{\text{BEATS}(x,z)}} \end{array}}$$
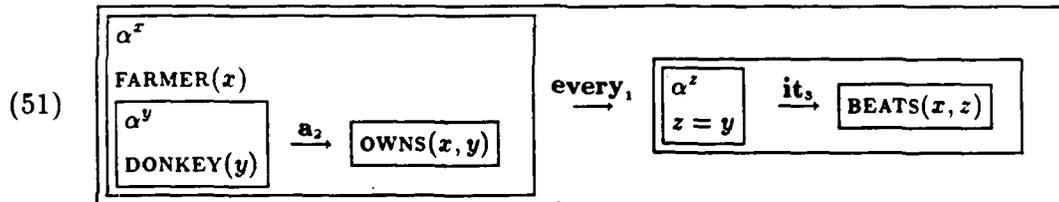
6.2 The Semantics

A model M for DRT$_1$ is a pair $\langle U, F \rangle$: F is an interpretation function, and U is a Link-type boolean algebra $< D, A, \leq, \sqcup >$ in which $\leq$ models the inclusion relation, A is the set of *atoms* of D, i.e., those elements of D such that no other element is included in them, and $\sqcup$ models Link's sum (that is, $x \sqcup y$ is the minimal element of U which includes both x and y).

The denotation of a DRT$_1$ expression with respect to M is defined as follows. The denotation of the basic conditions and of the marker introducers is the same as in DRT$_0$. The new atomic conditions have the following denotations:

1. $\|x \neq y\|^M = \{\langle f, f \rangle | \; f(x) \neq f(y)\}$

2. $\|\text{ATOM}(x)\|^M = \{\langle f, f \rangle | \; f(x) \in A\}$

3. $\|\text{GROUP}(x)\|^M = \{\langle f, f \rangle | \; f(x) \in U - A\}$

4. $\|x_i \prec x_j\|^M = \emptyset$ if $i > j$, $\{\langle f, f \rangle | \; x_i, x_j \in \text{DOM}(f)\}$ otherwise.

The denotation of the 'basic' DRS changes, since accessibility is now controlled by discourse structure and the type of NP. Rather than pairs $\langle f, f \rangle$, a basic DRS will denote pairs $\langle f, g \rangle$ where g contains additional values for the markers introduced by the DRS.

31

1. $\left\| \begin{array}{|c|} \hline \alpha^{x_1}, \ldots, \alpha^{x_n} \\ C_1 \\ \vdots \\ C_m \\ \hline \end{array} \right\|^M = \{\langle f, g\rangle|\ f \subset f_1 \subset \ldots \subset f_n \subset g_1 \subseteq \ldots \subseteq g_{m-1} \subseteq g,$
$\langle f, f_1\rangle \in \|\alpha^{x_1}\|^M$ and $\ldots \langle f_{n-1}, f_n\rangle \in \|\alpha^{x_n}\|^M$ and
$\langle f_n, g_1\rangle \in \|C_1\|^M, \ldots \langle g_{m-1}, g\rangle\|C_m\|^M\}$

The semantics of quantifier and referential DRSs will be a straightforward extension of the semantics of the conditional DRS in DRT$_0$. Let us first introduce the following definitions. Let $K_1 \xrightarrow{d_i} K_2$ be a quantifier or referential DRS, with determiner d, index i, and main marker x. Let f be an embedding such that $x \notin \text{DOM}(f)$. The set of DRS-*Satisfying Embeddings*, DSE, is defined as follows:

$$\text{DSE}(f, K) = \{h|\ \langle f, h\rangle \in \|K\|^M\}$$

and extended as follows to sets of embeddings:

$$\text{DSE}(\{f_1, \ldots, f_n\}, K) = \{h|\ \langle f_i, h\rangle \in \|K\|^M \text{ for some } i \in 1 \ldots n\}$$

I need now to make the notion of 'distancing' introduced in section 5.1 more precise. As said there, distancing has the effect of 'changing the perspective' on a certain situation, that is, making available 'outside' the DRS only the sets of participants to a situation 'as a whole', so that whereas before the closure the markers are available for singular anaphora, after they are only available for plural anaphora.

Definition 6.1 *Let E be a set of embedding pairs $\{\langle f, g\rangle\}$. DISTANCING(E) is the set of embedding pairs $E' = \{\langle f, h\rangle\}$ defined as follows. Let $x_i \ldots x_j$ be the markers over which g extends f. Then $\langle f, h\rangle$ will be in DISTANCING(E), where h is the embedding which is like g for the markers up to x_{i-1}, and then for $l = i \ldots j$, $h(x_l) = h_1(x_l) \sqcup \ldots h_k(x_l) \ldots$, for any h_k s.t. $\langle f, h_k\rangle \in E$.[12]*

We can now specify the denotations of quantified and referential DRS's:

1. $\|K_1 \xrightarrow{\text{every}_i} K_2\|^M = \text{DISTANCING}(\{\ \langle f, g\rangle, \text{where } f \text{ is an embedding such that for every } h \in \text{DSE}(f, K_1) \text{ there is an } h' \text{ s.t. } h' \in \text{DSE}(h, K_2), \text{ and } g \in \text{DSE}(\text{DSE}(f, K_1), K_2)\})$

2. $\|K_1 \xrightarrow{\text{most}_i} K_2\|^M = \text{DISTANCING}(\{\ \langle f, g\rangle|\ \text{where } f \text{ is an embedding such that for most } h \in \text{DSE}(f, K_1) \text{ there is an } h' \text{ s.t. } h' \in \text{DSE}(h, K_2), \text{ and } g \in \text{DSE}(\text{DSE}(f, K_1), K_2)\})$[13]

3. $\|K_1 \xrightarrow{a_i} K_2\|^M = \{\ \langle f, g\rangle|\ \text{DSE}(\text{DSE}(f, K_1), K_2) \neq \emptyset \text{ and } g \in \text{DSE}(\text{DSE}(f, K_1), K_2)\}$

---

[12]This semantics is correct for the examples in the previous section, but also predicts that in texts like *Every man who owns a donkey beats it. The donkeys hate their owner.*, *The donkeys* would refer to the set of all donkeys beaten by any man. To fix this problem it would be necessary to modify the defintion of distancing to allow for ambiguities.

[13]This definition suffers from the so called *proportion problem* [Heim, 1982; Rooth, 1987]. Let us consider a model with 100 farmers, 99 of which own a donkey and don't thrive whereas one, Pedro, owns 1000 donkeys and thrives. The definition of the semantics of $K_1 \xrightarrow{\text{most}_i} K_2$ above would predict that this model would make *Most farmers who own a donkey thrive* true, which is rather counterintuitive. The reason is that if the restriction is a DRS K like

32

4. $\|K_1 \overset{\text{he}_i}{\to} K_2\|^M = \{ \langle f, g \rangle |\ f(x) = f(y)$ for some marker y and $\langle f, g \rangle \in \|K_2\|^M \}$

(Analogous rules could be developed for $K_1 \overset{\text{she}_i}{\to} K_2$ and $K_1 \overset{\text{they}_i}{\to} K_2$, of course.) Next, the denotations of the negated DRS and the conditional, almost identical to those in DRT_0:

1. $\|\neg_i K\|^M = \{ \langle f, f \rangle |\ $ there is no g such that $\langle f, g \rangle \in \|K\|^M \}$

2. $\| K_1 \overset{\text{if-then}_i}{\to} K_2\|^M = \{ \langle f, f \rangle |\ $ for every h s.t. $\langle f, h \rangle \in \|K_1\|^M$ there is an l s.t. $\langle h, l \rangle \in \|K_2\|^M$, and g is one such $l \}$

(Note that, because of these definitions and of the definition of $x \prec y$, negation and conditionals still 'block' anaphoric reference in the same way that they do in DRT.) Finally, in order to define the semantics of scope forests, we have to specify more precisely the set of paths associated to them. This definition will be given in terms of an *extraction procedure*.

Definition 6.2 *An extraction procedure is an algorithm which 'pulls out' all the operators from a scope forest $\varphi[i_1 \ldots i_n]OC$, one after the other, and returns a* disambiguated *DRS (a DRS that is not a scope forest), under the constraint that if the ordering constraint $i < j$ is in OC, then the operator with index i is extracted before the operator with index j.*

What the 'extraction rules' do is fairly obvious and can be compared to what Quantifier Raising does. I will just give two examples:

negation:
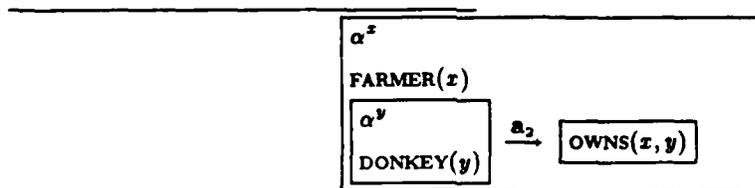
DSE(f,K) will include a different embedding for each *pair* of values for the markers x and y, and the fact that Pedro owns 1000 donkeys will make him a participant in 1000 pairs. The source of the problem, in short, is that the pairs are counted, rather than the number of possible distinct values that may be given to x. Rooth presented a solution to the problem which could easily be accomodated in my framework. The idea is to partition a set of embeddings into equivalence classes according to the value that the embeddings associate to x:

> EP(E,x), where E is a set of embeddings, and x is a marker, $= \{ E_a |\ $ each $h \in E_a$ is a member of E, and for all $h \in E_a\ h(x) = a$ for some $a \in U \}$

and then replace the definition of $K_1 \overset{\text{most}_i}{\to} K_2$ above with the one below:

- $\|K_1 \overset{\text{most}_i}{\to} K_2 \|^M = \{ \langle f, g \rangle |\ $ for most $E \in $ EP(DSE(f, K_1), x) there is an $h \in E$ s.t. $h' \in $ DSE(h, K_2) for some $h' \ldots \}$ (g is defined as before)

This definition does not suffer from the problem above, but it does not explain why the semantics of *most* should be defined in terms of equivalence classes, while the semantics of *every* should be based on pairs. A more illuminating solution would be to assume an ambiguity, originated by the possibility of choosing different farmers-owning-donkeys 'cases' in the restriction of a quantifier; developing this solution in detail, however, would go outside the scope of this paper.

33

indefinites:

$$\varphi \boxed{T_1 \; \text{R} \; (\; \mathbf{a}_i \; \boxed{\begin{array}{l} \alpha^x \\ \text{P}(x) \end{array}} \;)} \qquad \Rightarrow \qquad \boxed{\begin{array}{l} \alpha^x \\ \text{P}(x) \end{array}} \xrightarrow{\mathbf{a}_i} \boxed{T_1 \; \text{R} \; x}$$

The first extraction rule extracts a negation operator from a scope forest, and produces as a result a negated DRS with the scope forest from which the operator has been extracted in the scope of the negation operator. T_1 and T_2 can be either unextracted terms or markers. The second rule extracts an indefinite term from a scope forest, and produces a referential DRS in whose scope there is a new scope forest in which the marker x has replaced the indefinite term. It is worth observing that nothing prevents having more than one extraction rule per determiner, or extraction rules which operate on the scope forest as a whole (this will take care of the additional readings of examples like *Two examiners marked six scripts*). An 'empty extraction rule' will replace a scope forest without operators left with a basic DRS.

Definition 6.3 *Let K be the scope forest $\varphi[i_1 \ldots i_n]OC$ with indices $i_1 \ldots i_n$ and ordering constraints OC. The **paths associated to** SF are all the disambiguated DRS's K' which can be derived from K by an extraction procedure.* .

We can now define the denotation of a scope forest as follows:

1. $\| \; K : \varphi[i_1 \ldots i_n]OC \; \|^M = \bigcup \|P\|^M$, for every path P associated to K

6.3 The Inference Rules

The set of inference rules of DRT$_1$ includes, in addition to the traditional rules (Modus Ponens, Universal Instantiation, etc.) all the rules defined in section 5.1 and 5.4. Because of the change in the semantics of the basic DRS, we need to modify the definition of logical entailment given in section 4.2 as follows: if K_1 and K_2 are formulas, $K_1 \models K_2$ iff for all models M in which K_1 is true, for every pair $\langle f, g \rangle \in \|K_1\|^M$, there is an embedding h, $g \subseteq h$, such that $\langle f, h \rangle \in \|K_2\|^M$. With this new definition, the following theorem holds:

Theorem 6.1 *MP, UI, ROR, RAOQ, QORG, SFE, WUI, PA and ASTR are sound DRT$_1$ inference rules.*

Proof: I will show how one can prove this for MP; the proof for the other inference rules is analogous. Let $\langle f, g \rangle$ be a pair in the denotation of a formula K containing the two conditions P and $K_1 \xrightarrow{\text{if-then}_i} K_2$, where K_1 consists of the condition P and K_2 consists of the condition Q. Without loss of generality, we can assume that the two conditions are the 'last' conditions in K. Let f_n be an extension of f such that $\langle f_n, f_n \rangle \in \|K_1 \xrightarrow{\text{if-then}_i} K_2\|$ and $\langle f_n, g \rangle \in \|P\|$. Because of the definition of the conditional DRS, there must be an h such that $\langle g, h \rangle \in \|Q\|$. But then, $\langle f, h \rangle$ will be in the denotation of the formula K' obtained by adding Q at the end of K, QED.

7 Discussion

It is useful at this point to discuss the implications of the scope forest idea, and to state more clearly the position I am assuming with respect to certain topics in ambiguity processing which have been the object of much discussion.

7.1 Preferences

I have observed in section 2 that Kurtzman and MacDonald, and before them Ioup, answered in the affirmative to the question: Do people actually have preferences? The existence of preferences brings out other questions, however: (i) Do people apply these principles/preferences all the time? (ii) Which principles/preferences are actually used? (iii) Is it possible to derive them from some more general consideration? (iv) How can these preference rules be formalized?

I argued in section 2 that the 'scope forest theory' is compatible both with the 'lazy' model of disambiguation proposed in section 1 and with a 'fanatical' model of disambiguation in which the system always tries to come out with an interpretation. An intriguing hypothesis about question (i) is that whether people try to disambiguate or not depends on (a) how high the cost of disambiguating is and (b) how important it is for them to disambiguate. Kurtzman and MacDonald only studied sentences with two quantified NP's, and the determiners that appeared in those sentences were *every, each, the, the same, a* and *a different*. These quantifiers are 'easy to compute' – they can in fact be computed by DFAs [van Benthem, 1987]. It would be interesting to repeat the experiment with determiners which are 'harder' than the determiners above, such as *most* and *few*: for example, to try sentences like *Most male students date few undergrads*. Another source of complexity is the number of possible interpretations: again, the hypothesis would predict that as the number of interpretations gets larger, people would disambiguate less. This second prediction could be tested by running experiments on sentences containing numerical quantifiers, such as *Two examiners marked six transcripts.*.

It is also possible that people do not disambiguate in certain cases because they have no reason to do so. Hirst argues convincingly in [Hirst, 1990] for a similar strategy for systems processing; but there also seem to be linguistic reasons to believe this. Consider the following sentence:

(52) For simplicity of exposition, I will assume in most of the discussion
that variable names are magically correct from the start.   [Hirst, 1987]

The author is describing a program called Absity. Depending on how *the start* is interpreted (the start of the program, or the moment at which a variable is created), one can interpret it to take scope over *variable names*, or viceversa. Both interpretations are equally acceptable: deciding the scope of *the start* is not required for the understanding of (52).

Further evidence for the claim that we humans disambiguate only when we have some reason to do so can be found by observing how people refer: for example, how they use the demonstrative *that*. Asher discusses the 'sloppiness' exhibited by writers when using

that to refer to actions or plans in [Asher, 1990]. We have found that something similar also occurs when *that* is used to refer to objects. For example, in (53), *that* – if interpreted literally – could refer either to the engine, or to the boxcar, or to both:

(53) We're gonna hook up engine E2 to the boxcar at Elmira, and send that off to Corning.

disambiguating that reference is not, however, necessary to understand (53) – any disambiguation will probably lead to the same action on the part of the system which receives this order, since the boxcar and the engine are 'hooked up'[14]

Questions (ii) and (iii) – What are these principles? Is it possible to derive them from some more general principle? – are also interesting because answering them might help in finding some logic behind the disparate set of *Preference Heuristics* that have been proposed for NLP systems [VanLehn, 1978; Grosz *et al.*, 1987; Hurum, 1988a; Moran, 1988]. Moran lists in [Moran, 1988] eight preference heuristics. Of these, two can be related to the Quantifier Hierarchy Principle (3.1 - wide scope preference for *each* - and 3.2 - wide scope preference for WH-phrases but not over *each*), three establish a similar hierarchy between quantifiers, negation and modals (5.1 - strong preference for *some* to outscope negation; 5.2 - preference for negation to outscope *every*; and 5.3 - preference for *any* to outscope modals); and two are the Scope Constraint (presented in section 3) and May's generalization that quantifiers cannot be raised out of a relative clause. (Hurum [1988b] presents additional heuristics.)

I am afraid that an answer to questions (i)–(iii) won't be had for some time; answering question (iv), however, is a preliminary step, and can give some interesting insights on the more complicated problems. Giving a semantics to unscoped logical forms does not rule out scope forest reduction rules based on preferences. These preferences should be formalized as defeasible rules, however, with consequent problems in formalization. The hardest problems are how to cumulate evidence and how to choose among competing hypothesis. The abduction scheme of Hobbs and Stickel [Hobbs *et al.*, 1988] is an obvious possibility; another possibility is to use one of the probabilistic logics recently proposed. Once this extended logic is developed, one could attempt to characterize the cases of perceived ambiguity as follows: one recognizes an ambiguity whenever the preference rules are used (that is, there are reasons to disambiguate) and more than one plausible, but competing hypothesis are produced. One does not perceive the ambiguity if either the preference rules are not used, or they do not produce plausible disambiguation hypotheses.

7.2 The Role of Syntactic Constraints

One should not read in what I have written so far a claim that syntactic constraints play *no* role in determining scope relations. There is evidence to the contrary: May [1985], for example, shows that sentences like *Some student admires every professor* may be ambiguous in isolation, but in a VP-deletion context, like (54), the ambiguity evaporates.

(54) Some student admires every professor, but John doesn't.

---

[14]This sentence has been excerpt from the transcript of a conversation recorded at the University of Rochester for the TRAINS project.

It has also been brought to my attention by Anil Belvadi (p.c.) that in the Kanada language several cases of scope ambiguities are lexically resolved, that is, in sentences like *Every male student dates an undergrad* one would use a different determiner when *an undergrad* takes wide and narrow scope.

That syntactic factors can be used to disambiguate is not a problem for the theory I have been proposing. All that I am proposing is a method for resolving certain ambiguities using semantic and pragmatic knowledge; if the ambiguity can be resolved syntactically, the translation procedure will not produce a scope forest, but the appropriate disambiguated DRS. My claim is much weaker than the strong claim of Gawron and Peters in [Gawron and Peters, 1990](cfr. pages 53–56), that the determination of the scope of NP's is *entirely* determined by contextual factors (Gawron and Peters call them *circumstances*). In order to support this stronger claim one should give some sort of explanation for data like (54).

7.3 How Many Interpretations are Computed?

Another question left open by Kurtzman and MacDonald's experiments is whether people, when they do disambiguate, do indeed compute only one interpretation, or perhaps a couple of them - as predicted by the theory presented in this paper, and as one would expect on the basis of 'common sense' arguments like those presented in section 1 – or compute all interpretations in parallel, as predicted for example by the disjunction theory. Kurtzman and MacDonald's data suggest that people compute both interpretations of sentences like *Every boy climbed a tree* – they claim that people always initially compute both interpretations and then usually quickly adopt one interpretation and delete the other [Kurtzman, p.c.]. Once more, however, running tests with sentences with a large number of interpretations would be the only way of discriminating among the competing hypotheses. The sheer number of interpretations should make it possible to verify whether indeed people compute them all and then 'filter', or rather compute a few (possibly using preference heuristics).

7.4 Other Types of Ambiguity?

One may also wonder if this kind of semantics can be used to handle different kinds of ambiguities. The reasoning goes as follows. When given a sentence with, say, a lexical ambiguity, the translation procedure will produce more than one logical form, or, alternatively, a logical form containing some construct for representing the ambiguity 'in loco', like the *Polaroid words* proposed by Hirst [Hirst, 1987]. The second kind of representation would be, of course, more in keeping with the scope forest proposal, but in both cases we should be able to assign as a denotation to the logical form a set of pairs (if more than one logical form is produced by the translation, this set would be the union of the sets denoted by each of these logical forms). It would be necessary to define disambiguation rules and maybe preference rules which do the job of spreading activation in Hirst's system, of course, but at least in principle, this approach seems practicable. Other kinds of semantic/pragmatic discourse disambiguation processes – for example, discourse structure disambiguation – seem amenable to this kind of treatment as well. What this means is that we would be able to assign a denotation to a disambiguated parse tree, and therefore all of the semantic and

pragmatic processing of the system could be done 'deductively': the system could just be fed a parse tree.

If we were able to do this, we would have a principled way to formalize the idea that discourse disambiguation does not proceed sequentially - lexical disambiguation first, then scope disambiguation, then tense interpretation, then reference disambiguation, and so forth - but instead as a form of constraint propagation process, in the sense that one starts with a set of constraints which are compatible with a large number of possible interpretations, and then proceeds to reduce the number of allowed interpretations by inferring new constraints from the existing ones. The idea that discourse interpretation processes are interrelated in this way has been made popular by, among others, J. Hobbs [Hobbs, 1979; Hobbs *et al.*, 1988], and methods for implementing this kind of disambiguation process have been proposed, besides Hobbs, by K. Dahlgren [Dahlgren, 1988], by Charniak and Goldman [Charniak and Goldman, 1988; Goldman, 1990], and by the University of Rochester group working on the Discourse System project [Allen *et al.*, 1989]. On the psycholinguistic side, a model of processing in which different kinds of discourse information are related has been proposed by Crain and Steedman [Crain and Steedman, 1985]. None of these projects was focused on the problem of ambiguity, however, so the criticism to 'traditional' systems applies to each of them - in all of them, the only way to give a semantics to an ambiguous statement is to computing all the interpretations and then either choose one or use a disjunctive statement.

One might even carry the hypothesis further, and claim that even structural ambiguities like prepositional attachment could be handled in this way. That is, one could feed the semantic interpreter a 'compact' representation for an ambiguous parse tree of the kind proposed by Seo and Simmons [Seo and Simmons, 1989], instead of a disambiguated parse tree. The denotation of this object would be, once more, the union of the denotations of the unambiguous parse trees. Serious work on the idea is clearly needed before its usefulness can be judged.

8 Conclusions and Future Work

The main aim of this paper was to show that a natural language processing system can perform inferences on the basis of what it has been told without doing scope disambiguation. A simple parser-with-reasoner system built according these ideas has been implemented, and the implementation of a second version is under way. This new version will be used as the reference and scope disambiguation module of the discourse system TRAINS [Allen and Schubert, 1991]. My main goals for the foreseeable future are (i) to explore the properties of DRT_1 in more detail, looking for counterintuitive properties[15]; (ii) to extend the treatment to cover other operators which create scope ambiguities, most notably tense operators, and (iii) to develop a version of the logic in which I can formalize preference rules, and compare the two strategies – leaving the ambiguities in place, vs. generating one (or few) interpretations using preferences.

8.1 Reasoning Without Disambiguating?

The first direction of research is to discover what kind of inferences one can do with a non-disambiguated logic - that is, how the set of 'weak' inference rules proposed in section 5 can be extended to a complete set of inference rules, and how this set will look like. The ultimate goal is to develop a logic in which one *never* disambiguates – that is, one in which only scope forests are used, and the usual 'scoped' forms of representation are only used as shorthands.

A crucial question is whether we get any actual improvement in performance by leaving the ambiguities in place, that is, whether inferencing with DRT_1 is so expensive as to offset the advantages gained from not having to produce all the interpretations. A general answer would require determining the complexity of reasoning with DRT_1. My first objective will be to perform an experimental study, by implementing a new version of the scoping and reference disambiguation module of TRAINS and comparing its performance both with the existing module and with a module which works according to the proposals of Moran and Hurum.

8.2 Formalizing Preferences

To formalize the kind of preferences discussed by Kurtzman and MacDonald, as well as to represent the disambiguating information originated from lexical and world knowledge, I will need a logic in which it is possible to jump to conclusion on the basis of uncertain knowledge, and to choose between competing possibilities.

For example, it is a lexical property of the predicate *comes with* that the argument in subject position takes scope over the argument in object position, so that *Every chess set comes with a spare pawn* has no reading in which *a spare pawn* scopes over *Every chess set*.

---

[15]Because of the use of partial situations and the formulation of the definition of the marker introducers it doesn't seem to suffer from some of the unwanted properties of, e.g., certain dynamic logics: for example, $A \rightarrow A$ is a logical axiom.

In order to express such 'defeasible disambiguation axioms' a **Disambiguation Schema** can be used, whose general form would be

$$S \longrightarrow \{i<j\}$$

where S is the scope of the quantifier with the strictest scope and i and j are indices of operators. The disambiguation schema for the predicate *comes with*, for example, would be

$$[i \ comes\text{-}with \ j] \longrightarrow \{i < j\}.$$

This schema is an abbreviation for a list of axiom schemata each of which adds the ordering constraint $i < j$ to a scope forest. Unfortunately, schemata of this kind do not always hold: they can only be used if the logic can express some form of uncertainty.

A number of proposals for formalizing the process of choosing among preferences have been made, among which belief nets [Pearl, 1988] and logics in which it is possible to order the propositions in the knowledge base according to some principle [Nebel, 1990]. It's not yet clear whether any of these proposals is appropriate for my purposes.

8.3 Tense and Scope Ambiguities

The parser used in TRAINS produces a logical form in which not only quantified NP's, but also certain operators are unscoped, among which negation and tense operators[16]. A scoping algorithm is then applied, and then the logical form is translated into Episodic Logic, the logic developed by Len Schubert and Chung-Hee Hwang [Schubert and Hwang, 1990] by a final step in which tense operators are eliminated and replaced by formulas which contain *episodic variables* using an additional contextual structure, *Tense Trees* [Schubert and Hwang, 1990].

In order to test whether the scope forest idea can be used in practice to represent scope ambiguities other than those between NP's, I plan to study how to integrate the tense tree interpretation mechanism with the other scope ambiguity resolution processes, so that it won't be necessary to decide upon a scoping before applying the tense interpretation rules.

8.4 Other Issues

A question that needs to be addressed in more detail is the relation between scope ambiguities and, for example, collective/distributive ambiguities, in order to see better the relation between this approach and, for example, the approach proposed by Bunt. The reference interpretation model also needs to be extended. I'm in particular interested in studying sentences like *Every chess set comes with a spare pawn. It's taped to the bottom.*, in which the interface between reference and scope disambiguation seems totally different from the one presented in the examples in section 5.4.

---

[16]The logical form is essentially that proposed in [Schubert and Pelletier, 1982].

Acknowledgements

References

[Allen, 1991] J. Allen, "Natural Language, Knowledge Representation, and Logical Form," Technical Report 367, University of Rochester, Rochester, NY, January 1991.

[Allen et al., 1989] J. Allen, S. Guez, L. Hoebel, E. Hinkelman, K. Jackson, A. Kyburg, and D. Traum, "The Discourse System Project," Technical Report 317, University of Rochester, Rochester, NY, November 1989.

[Allen and Schubert, 1991] J. Allen and L.K. Schubert, "The TRAINS Project," TRAINS Technical Note 91-1, University of Rochester, Department of Computer Science, 1991.

[Allen, 1987] J. Allen, *Natural Language Understanding*, Benjamin Cummings, Menlo Park, CA, 1987.

[Alshawi et al., 1988] H. Alshawi, D. M. Carter, J. van Eijck, R.C. Moore, D. B. Moran, F.C.Pereira, and A.G.Smith, "Interim Report on the SRI Core Language Engine," Technical Report CCSRC-5, SRI International, Cambridge, Cambridge, England, 1988.

[Alshawi and van Eijck, 1989] H. Alshawi and J. van Eijck, "Logical Forms in the Core Language Engine," In *Proc. ACL-89*, Vancouver, CA, June 1989.

[Asher, 1990] N. Asher, "Abstract Objects, Semantics and Anaphora," Unpublished manuscript, 1990.

[Barwise, 1987] J. Barwise, "Noun Phrases, Generalized Quantifiers and Anaphora," In Gärdenfors [Gärdenfors, 1987], pages 1–30.

[Barwise and Perry, 1983] J. Barwise and R. Perry, *Situations and Attitudes*, The MIT Press, 1983.

[Bunt, 1985] H. Bunt, *Mass Terms and Model Theoretic Semantics*, D. Reidel, 1985.

[Charniak and Goldman, 1988] E. Charniak and R. P. Goldman, "A Logic for Semantic Interpretation," In *Proc. ACL-88*, Buffalo, NY, June 1988, pages 87–94.

[Chierchia and McConnell-Ginet, 1990] G. Chierchia and S. McConnell-Ginet, *Meaning and Grammar: An Introduction to Semantics*, The MIT Press, 1990.

[Crain and Steedman, 1985] S. Crain and M. Steedman, "On not being led up the garden path: the use of context by the psychological syntax processor," In D. R. Dowty, L. Karttunen, and A. M. Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical perspectives*, pages 320–358. Cambridge University Press, New York, 1985.

[Dahlgren, 1988] K. Dahlgren, *Naive Semantics for Natural Language Understanding*, Kluwer, Norwell, MA, 1988.

[Fodor, 1982] J. D. Fodor, "The Mental Representation of Quantifiers," In S. Peters and E. Saarinen, editors, *Processes, Beliefs and Questions*. D. Reidel, Dordrecht, 1982.

[Gärdenfors, 1987] P. Gärdenfors, editor, *Generalized Quantifiers*, D. Reidel, Dordrecht, The Netherlands, 1987.

[Gawron and Peters, 1990] J. M. Gawron and S. Peters, *Anaphora and Quantification in Situation Semantics*, volume 19 of *Lecture Notes*, CSLI, 1990.

[Goldman, 1990] R. P. Goldman, *A Probabilistic Approach to Language Understanding*, PhD thesis, Brown University, Department of Computer Science, Providence, RI, 1990.

[Groenendijk and Stokhof, 1990] J.A.G. Groenendijk and M.J.B. Stokhof, "Dynamic Montague Grammar," In L. Kálmán and L. Pólos, editors, *Papers from the 2nd Symposium on Logic and Language*. Akadémiái Kiadó, Budapest, 1990.

[Grosz et al., 1987] B. Grosz, D. Appelt, P. Martin, and F. Pereira, "TEAM: An experiment in the design of transportable natural language interfaces," *Artificial Intelligence*, 32(2):173–244, 1987.

[Heim, 1982] I. Heim, *The Semantics of Definite and Indefinite Noun Phrases*, PhD thesis, University of Massachusetts at Amherst, 1982.

[Hirst, 1987] G. Hirst, *Semantic Interpretation and the resolution of Ambiguity*, Studies in Natural Language Processing. Cambridge University Press, Cambridge, UK, 1987.

[Hirst, 1990] G. Hirst, "Mixed-depth representation for natural language text," In *Proc. AAAI Spring Symposium*, pages 10–14, 1990.

[Hobbs et al., 1988] J. R. Hobbs, M. Stickel, P. Martin, and D. Edwards, "Interpretation as Abduction," In *Proc. ACL-88*, Buffalo, NY, June 1988, pages 95–103.

[Hobbs, 1979] J. Hobbs, "Coherence and Coreference," *Cognitive Science*, 3:67–90, 1979.

[Hobbs, 1983] J. Hobbs, "An Improper Treatment of Quantification in Ordinary English," In *Proc. ACL-83*, pages 57–63, Cambridge, MA, June 1983.

[Hobbs and Shieber, 1987] J. Hobbs and S. Shieber, "An Algorithm for Generating Quantifier Scopings," *Computational Linguistics*, 13(1-2):47–63, January-June 1987.

[Hurum, 1988a] S. Hurum, "Handling Scope Ambiguities in English," In *Proc. Second Conference on Applied Natural Language Processing*, 1988.

[Hurum, 1988b] S. Hurum, "Handling Scope Ambiguities Using Domain-independent Heuristics," Technical Report TR 88-12, University of Alberta, Edmonton, Canada, June 1988.

[Ioup, 1975] G. Ioup, "Some Universals for Quantifier Scope," In J. Kimball, editor, *Syntax and Semantics 4*, pages 37–58. Academic Press, New York, 1975.

[Johnson-Laird, 1977] P. Johnson-Laird, "Reasoning with Quantifiers," In P. Johnson-Laird and P.C. Wasow, editors, *Thinking*. Cambridge University Press, Cambridge, MA, 1977.

[Kamp, 1981] H. Kamp, "A Theory of Truth and Semantic Representation," In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Formal Methods in the Study of Language*. Mathematical Centre, Amsterdam, 1981.

[Kamp, 1988] H. Kamp, "Some Remarks About the semantic Representation of certain Plural Constructions of English, and Suggestions for a Procedural Treatment of Plural Pronoun Anaphora," Contribution to Acord Workshop, October 1988.

[Kempson and Cormack, 1981] R. Kempson and A. Cormack, "Ambiguity and Quantification," *Linguistics and Philosophy*, 4(2):259–310, 1981.

[Kurtzman and MacDonald, 1991] H. S. Kurtzman and M. C. MacDonald, "Interpretation Preferences for Quantifier Scope Ambiguities," In *Proc. SALT-1*, 1991.

[Lakoff, 1971] G. Lakoff, "Semantic Interpretation in Generative Grammar," In D. Steinberg and L. Jakobovits, editors, *Semantics: An interdisciplinary reader in philosophy, linguistics, anthropology, and psychology*. Cambridge University Press, 1971.

[Link, 1987] G. Link, "Generalized Quantifiers and Plurals," In Gärdenfors [Gärdenfors, 1987], pages 151–180.

[May, 1985] R. May, *Logical Form in Natural Language*, The MIT Press, 1985.

[Moran, 1988] D. Moran, "Quantifier Scoping in the SRI Core Language Engine," In *Proc. ACL-88*, Buffalo, NY, June 1988, pages 33–40.

[Nebel, 1990] B. Nebel, *Reasoning and Revision in Hybrid Representation Systems*, Springer-Verlag, 1990.

[Partee, 1975] B. Partee, "Comments on C.J.Fillmore and N.Chomsky's papers," In D. Austerlitz, editor, *The scope of American linguistics*. De Ridder Press, The Netherlands, 1975.

[Pearl, 1988] J. Pearl, *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[Poesio and Zucchi, 1991] M. Poesio and S. Zucchi, "On Telescoping," In Preparation, 1991.

[Reinhart, 1983] T. Reinhart, *Anaphora and semantic interpretation.*, Croom Helm, London, 1983.

[Roberts, 1987] C. Roberts, *Modal Subordination, Anaphora, and Distributivity*, PhD thesis, University of Massachusetts at Amherst, 1987.

[Rooth, 1987] M. Rooth, "Noun Phrase Interpretation in Montague Grammar, File Change Semantics, and Situation Semantics," In Gärdenfors [Gärdenfors, 1987], pages 237–268.

[Schubert and Hwang, 1990] L. K. Schubert and C. H. Hwang, "An Episodic Knowledge Representation for Narrative Texts," Technical Report 345, University of Rochester, Rochester, NY, 1990.

44

[Schubert and Pelletier, 1982] L. K. Schubert and F. J. Pelletier, "From English to Logic: Context-Free Computation of 'Conventional' Logical Translations," *American Journal of Computational Linguistics*, 10:165–176, 1982.

[Schubert and Pelletier, 1988] L. K. Schubert and F. J. Pelletier, "Generically Speaking," In G. Chierchia, B. H. Partee, and R. Turner, editors, *Properties, Types and Meaning, vol. II*, pages 193–268. Kluwer, Dordrecht, 1988.

[Seo and Simmons, 1989] J. Seo and R. Simmons, "Syntactic graphs: A representation for the union of all ambiguous parse trees," *Computational Linguistics*, 15(1):19–32, 1989.

[van Benthem, 1987] J. van Benthem, "Towards a Computational Semantics," In Gärdenfors [Gärdenfors, 1987], pages 31–72.

[VanLehn, 1978] K. A. VanLehn, "Determining the Scope of English Quantifiers," Technical Report AI-TR-483, Artificial Intelligence Laboratory, MIT, Cambridge, MA, 1978.

[Webber, 1978] B. Lynn Webber, "A Formal Approach to Discourse Anaphora," Report 3761, BBN, Cambridge, MA, May 1978.