Quantification in natural language is an important phenomena that seems to touch on some pragmatic and inferential aspects of language understanding. In this paper we focus on quantifier scope ambiguity and suggest a cognitively plausible model that resolves a number of problems that have traditionally been addressed in isolation. Our claim here is that the problem of quantifier scope ambiguity can not be adequately addressed at the syntactic and semantic levels, but is an inferencing problem that must be addressed at the pragmatic and discourse levels.

Quantification in Natural Language

Quantification is an important phenomena that seems to touch on a number of inferential and pragmatic aspects of language understanding. To illustrate the "pragmatic" aspect of this problem, consider the following examples involving the problem of quantifier scope ambiguity:

1. *John advertised a restaurant on every street*
2. *John visited a restaurant on every street*

In compositional semantics in the Fregean tradition, e.g. (Montague, 1974), the form dictates the construction of a certain logical form (LF) thereby forcing a specific scope ordering of quantifiers. Having identical syntactic derivations, the same semantic rule would be used to construct a LF for (1) and (2), essentially making the interpretation of the noun phrase (NP) "a restaurant on every street" independent of the surrounding context.

Clearly, this is inadequate since, in the actual world we live in, a speaker of ordinary English might admit the possibility of a single restaurant being advertised on every street ("a" outscopes "every") but is not likely to admit a reading of (2) which implies the physical existence of the same restaurant on every street.

Recognizing this type of ambiguity, Montague devised a method by which syntactically unambiguous sentences could have multiple derivation trees. As noted by (Thomason 1974) and (Gamut 1991), the quantifying-in rule that Montague suggested can be seen as an "invention" to salvage compositional semantics. The rule is in fact a template, a meta rule, for an infinite number of derivations, each representing to a possible scope ordering. A more computationally tractable technique was later proposed by (Cooper 1983). "Cooper Storage" was essentially a method by which quantifier scope ordering commitments can be delayed until further information could be brought into light. This general technique in various forms has been developed and extended by various computational linguists, such as (Allen 1987), (Alshawi 1990), (Hobbs and Shieber 1987), and (Pereira 1989).

Computational Models of Quantification

The problem of quantifier scope ambiguity presented a challenge to compositional semantics at two levels. In compositional semantics there is a semantic rule corresponding to every syntactic rule (this is the rule-to-rule correspondence in Montague's PTQ). This tandem between syntax and semantics meant that quantifier scope ambiguity had to be translated into a "derivational ambiguity" (see Dowty et al. 1981), as Montague's invention of the quantifying-in rule had done.

Quantifier scope ambiguity also presented a challenge to compositional semantics at the pragmatic level. In particular, computational systems were still faced with the problem of deciding on the most plausible scope ordering. In computational linguistics a number of techniques have been suggested. Typically, a sentence is first translated into a scope-neutral (Allen 1987) or a quasi LF (Alshawi 1990) with no scope ordering commitments. A number of syntactic and semantic rules are then incrementally applied to reduce the number of possible readings. Computational models for incremental interpretation must be carefully crafted if one is to stay committed to compositionality. Elaborate models at such attempts are given in (Pereira and Pollack 1991) and (Harper 1992). These models do not directly address the problem of scope ambiguity however. Instead, they provide a computational framework for the resolution of quantifier scope (and other) ambiguities.

---

2 Derivation trees in the PTQ system can thus be seen as logical forms.
Regarding the resolution of quantifier scope ambiguity, several algorithms that apply a number of syntactically-motivated "preference rules" have been suggested. Instead of inferring a most plausible reading, these methods suggest a combinatorial approach where less likely readings are incrementally disallowed. See (Allen 1987), (Hobbs & Shieber 1987), (Moran 1988) and (Park 1995).

While a number of important results have been achieved, the problem of quantifier scope ambiguity is still largely unresolved. Ironically, while admitting multiple LFs (multiple derivations in PTQ) for a syntactically unambiguous sentence is a recognition of the semantic nature of the problem, existing approaches to scope ambiguity are for the most part syntactic in nature.

In our view the resolution of quantifier scope ambiguity is an inferencing problem that operates at the pragmatic level. In the remaining of this section we briefly discuss some of the existing problems that strongly suggest the need for a pragmatic and commonsense reasoning approach. In section 2 we introduce the notion of a "quantificational restriction" which is a piece of commonsense knowledge that we assume speakers of natural language have recourse to. In the remainder of the paper we use quantificational restriction in the resolution of scope ambiguity and show how the model solves some existing problems that relate to quantification.

Combinatorial Puzzle and Branching Quantifiers
The number of possible readings for a logical form (LF) with \( n \) quantifiers is \( n! \). An indefeasible approach by which various rules are incrementally applied to eliminate less likely readings suggests that speakers of ordinary English initially consider all 24 possible readings of the sentence the first to point out that quantifiers are not always linear as has been initially suggested by Frege. Instead, it was suggested that in (3), for example, there are two pairs of quantifiers that are independent in regard to scope. While it is clear that some quantifiers are branching, extensive studies by (Faccounier 1975), (Barwise 1979), (May 1989), and (Sher 1990), among others, have not settled the issue of "under what conditions do quantifiers branch?" In fact, (Gurevich 1988) has shown that the problem is NP hard, which suggests that a computationally tractable solution must rely on some "heuristics." In a later section we suggest an explanation to quantifier branching.

The Transformational Puzzle
One of the strongest indications of the pragmatic and inferential aspects of the quantification problem is the so-called "transformational puzzle." Consider the following:

(4) Ten million people sepak every Slavic language
(5) Every Slavic language is spoken by ten million people

While syntactic theories in the Chomskian tradition suggest that the active and passive forms are semantically equivalent, the meanings implied by (4) and (5) are radically different. While we are speaking of 10 million people in (4), (5) implies the existence of \( 10 \times n \) million people where \( n \) is the number of Slavic languages. Clearly, therefore, quantifier scope is a phenomena that must be operating beyond the syntactic and semantic levels.

Commonsense Reasoning
As (1) and (2) suggest, the resolution of quantifier scope ambiguity is a process that involves commonsense reasoning. To illustrate this point further, consider:

(6) \( \{ \text{All} \} \) MIT students submitted a paper to AAAi97
(7) \( \{ \text{Several} \} \) MIT students submitted a paper to AAAi97
(8) \( \{ \text{Two} \} \) MIT students submitted a paper to AAAi97

(6) through (8) differ only in the relative ratio of "MIT students" to "a paper submitted to AAAi-97." While it is highly unlikely for all MIT students to submit the same paper, this possibility increases when speaking of several or two students. We argue that the inferencing involved in these examples has to do with our commonsense knowledge of the submit relation between a student and a paper, namely that the relation is typically a few-to-1 relation. We call this specific knowledge the "quantificational restriction."

Quantificational Restriction
In this section we introduce quantificational restriction (QR). We claim that speakers of ordinary English associate...
with a binary relation \( R \) between two concepts \( C_1 \) and \( C_2 \) a quantificational restriction (QR), defined as:

\[
(9) \quad QR(R,C_1,C_2) = \langle m_1, m_2 \rangle
\]

where

\[ m_1 = \text{number of } C_1 \text{ that are in relation } R \text{ to the same } C_2 \]
\[ m_2 = \text{number of } C_2 \text{ that the same } C_1 \text{ could be related to by } R \]

To motivate the use of quantificational restrictions we discuss two examples. Consider the (physically located) *On* relation defined between a *house* and a *street*, and the *Eat* relation defined between a *child* and a *cookie*. We suggest that given an appropriately defined context, which we will shortly discuss, speakers of ordinary English must have a quantificational restriction defined on these relations. For the moment we assume that these values exist in some individual’s knowledge base as:

\[
(10) \quad QR(On, House, Street) = \langle \text{many,1} \rangle
\]
\[
(11) \quad QR(Eat, Child, Cookie) = \langle 1, \text{several} \rangle
\]

That is, we have assumed that some individual “believes” that typically there are many houses located on a given street (and that no house can be on more than one street), and that a single child could eat several cookies.

Assuming that such quantificational restrictions exist in the speaker’s knowledge base, one must also consider a number of other parameters that QR must depend on.

**Possible Worlds**

The QR in (10) reflects some individual belief that a house can be physically located on one street, where there could be many. Clearly, this belief is a function of a possible world, in the sense of Montague (1974), as it is conceivable that some individual would associate a different QR in some fictional world where a house might be located on several (or indeed all) streets.

In our model, therefore, it is not crucial for a QR to have any specific value, but that such QRs *do exist* in the knowledge base. That is, we are committed to the assumption that understanding quantification must involve the evaluation of a QR given a fully specified context, which must include a Montagovian index of interpretation.

**Modal and Temporal Aspects**

Our intention is to precisely define the notion of an “appropriately defined context” for the dynamic evaluation of a quantificational restriction. In addition to an index of interpretation we suggest that modal and temporal aspects effect the dynamic evaluation of QR.

**Modality.** Consider again the QR given in (10), with a similar example of the (physically located) *On* relation:

\[
(10) \quad QR(On, House, Street) = \langle \text{many,1} \rangle
\]
\[
(12) \quad QR(On, Book, Shelf) = \langle \text{many,1} \rangle
\]

As in (10), the QR in (12) reflects some individual’s belief that a book can be physically located on one shelf, where there could be many. However, there is an important difference between (10) and (12) that is not captured by the quantificational restriction as we have defined it so far. While it might be necessary for a house to be located on some street, a book not be on a shelf. This is a crucial distinction that appears at the linguistic level to effect the choice of quantifier scope. Consider the following:

\[
(13) \quad \text{John visited every house on a street}
\]
\[
(14) \quad \text{John placed every book on a shelf}
\]

While there is a tendency to assume a single street in (13) (by having “a” outscope “every”) the same does not seem to be the case in (14). Assuming it is necessary for a house to be located on some street, the direct reading of (13) does not add any significant information. Thus, on Gricean grounds one would reject the reading that implies “John visited every house – that happened to be – on a street.” The case in (14) is quite different, since there is quite a bit of information in the reading where “every” outscopes “a” (since a book might very well be placed on a desk.) This additional dimension, that of the modality of a relation, therefore, seems to explain quantifier scope at a much deeper level. Note that when “a” precedes “every” in the surface structure, as in the following:

\[
(15) \quad \text{John visited a house on every street}
\]
\[
(16) \quad \text{John placed a book on every shelf}
\]

The quantifier scope ordering must be reversed in both cases on QR grounds, since neither a house nor a book can physically occupy more than one space.

In this paper we will assume that a relation between two concepts can be necessary (Nec), typical (Typ), or simply possible (Pos). We further assume that \( \text{Nec} \preceq \text{Typ} \preceq \text{Pos} \).

**Temporal Aspect.** Consider the following QR defined on the relation *Lift* between *man* and *piano*:

\[
(17) \quad QR(Lift, Person, Man) = \langle 2',1 \rangle
\]

This QR reflects some individual belief that it takes at least two men to lift a single piano. Two men, however, can lift more than one piano at different points in time. This temporal aspect is crucial to quantification and has in fact

\[ \text{Typ} \preceq \text{Pos} \]

3 The similarity between “house-On-street” and “book-On-shelf” suggests that this sense of “On” defines a general schema. However, the two cases differ slightly with respect to modality suggesting a further classification. An extension to this work that we are currently investigating is to identify these general schemas in the spirit of (Lakoff 1987).

4 Assuming it is necessary for a house to be on one street, than it is typical, and trivially, possible. Note that the reverse reasoning is not valid.
been investigated extensively, most notably by (Verkuyl 1989). To illustrate, consider the following:

(18) Four men lifted two pianos to the second floor
(19) Two men lifted four pianos to the second floor

Given the QR in (17), the two possible scope orderings in (18) would be allowed. The reading that most readers would choose for (19), however, can only be explained by admitting at least two lifting events. As will be explained below, the QR in (17) not only aids in choosing a plausible reading but also defines all the possible ways the collective and distributive readings can occur.

In this paper we assume a relation between two concepts, when referring to an event, can refer to a single (T) or multiple (l) events. We further assume that T, M, and Q to be ordered sets. Assuming a monotone increasing order over these sets, implication between two QRs can now be defined as follows:

\[
\begin{align*}
QR(\text{Context},(R,C_1,C_2)) &= (m_1,m_2) \\
\text{where} & \\
\text{Context} &= (\text{Index},\$\text{Temp},\$\text{Mod}) \\
\$\text{Temp} &\in T = \{T,\bar{T}\} \\
\$\text{Mod} &\in M = \{\text{Pos,Typ,Typ}\} \\
(m_i &\in Q = \{\text{none,one,...,few,several,many,most,all}\}
\end{align*}
\]

We also assume T, M, and Q to be ordered sets. Assuming a monotone increasing order over these sets, implication between two QRs can now be defined as follows:

\[
\begin{align*}
&\text{let} \\
&qr_A = QR((\text{Index},t_A,m_A),(R,C_{A1},C_{A2})) = (n_1,n_2) \\
&qr_B = QR((\text{Index},t_B,m_B),(R,C_{B1},C_{B2})) = (n_1,n_2) \\
&\text{then} \\
&\forall t_A,t_B \in T, \forall m_A,m_B \in M \\
&[(q_1 < q_2) \Rightarrow (t_1 < t_2) \land (m_1 < m_2)]
\end{align*}
\]

Implication between two QRs defines a partial order of general relational schemas. Recalling examples (10) and (12) from above, the fully specified QRs that were assumed in these examples are as follows:

\[
\begin{align*}
qr_1 &= QR((\text{Index},\$\text{Temp},\text{Typ}),(\text{On,House,Street})) = (\text{many,1}) \\
qr_2 &= QR((\text{Index},\$\text{Temp},\text{Typ}),(\text{On,Book,Shelf})) = (\text{many,1})
\end{align*}
\]

Note that we have assumed that it is necessary for a house to be on a street, while it is only typical for a book to be on a shelf. While both QRs are \(q_1 < q_2\) since \(\text{Typ} < \text{Nec}\). (See footnote 2.)

**Inferences Using Quantificational Restriction**

In this section we briefly discuss the use of quantificational restriction in resolving scope ambiguities.

**Quantifier Scope Ambiguity**

In resolving quantifier scope ambiguity we assume a scope neutral logical form (LF) as in (Allen 1987) and (Alshawi 1987). In this paper we assume the following LF corresponding to a binary relation defined on two quantified term phrases:

\[
(22) R(q_1C_1,q_2C_2)
\]

Given a QR(Context,(R,C_1,C_2)) = \((n_1,n_2)\), we define the following quantifier scope rule:

\[
(23) \begin{align*}
\text{SCOPE}(R(q_1C_1,q_2C_2)) &= \begin{cases} \\
R(q_1C_1,q_2C_2) & \text{if} \ (n_1,n_2) \subseteq (q_1,n_2) \\
R(q_2C_2,q_1C_1) & \text{otherwise}
\end{cases}
\end{align*}
\]

where

\[
(q_1,n_2) \text{ is analogous to numerical multiplication; and}
(n_1,n_2) \subseteq (m_1,m_2) \iff ((n_1 \leq n_2) \Rightarrow (m_1 \leq m_2)) \land ((n_1 \geq n_2) \Rightarrow (m_1 \geq m_2))
\]

Note that the product of a quantifier and a numerical value is reduced to ordinary multiplication\(^6\). We now consider some examples with the following relevant QR's:

\[
\begin{align*}
QR((I,\$\text{Temp},\text{Typ}),(\text{Submit,Student,Paper})) &= (2^-1) \\
QR((I,\$\text{Temp},\text{Typ}),(\text{Lift,Person,Piano})) &= (2^+1) \\
QR((I,\$\text{Temp},\text{Nec}),(\text{On,House,Street})) &= (\text{many,1})
\end{align*}
\]

That is, we suggest that in some individual's knowledge base it is assumed that, typically two or less co-author a

\(^5\) We assume an index of interpretation referring to the actual world we live in. Temporal and modal aspects have default values, and thus the context is assumed to be fully specified.

\(^6\) A product involving at least one linguistic quantifier forces us to assign numerical values for the linguistic quantifiers (i.e., to "quantify" the linguistic quantifiers.) This is a lengthy discussion since the exact numerical value of a quantifier in a certain context is a function of the size of the quantified set. That is, "many people in the US" is much larger than "many students at MIT." In this paper we will not cover such situations, instead we will only consider obvious values.
paper, that typically it takes two or more to lift a piano, and that necessarily a house is located on a single street, where there could be many. Now consider the following:

(25) 2 men lifted every piano to the second floor
(26) 2 men lifted 3 pianos to the second floor
(27) 4 men lifted 2 pianos to the second floor
(28) John visited a house on every street
(29) many MIT students submitted a paper to AAAI - 97
(30) 2 MIT students submitted a paper to AAAI - 97

Using the QRs given in (24) the reader can easily verify that repeated applications of the SCOPE rule given in (23) yields the following:

(25') reverse order: <2+, 1> ⊈ <2 (1), every (2+)>
(26') reverse order: <2+, 1> ⊈ <2 (1), 3 (2+)>
(27') allow order: <2+, 1> ⊊ <4 (1), 2 (2+)>
(28') reverse order: <many, 1> ⊈ <a (1), every (many)>
(29') allow order: <2-, 1> ⊊ <many (1), a (2-)>
(30') reverse order: <2-, 1> ⊈ <2 (1), a (2-)>

The case in (27') is particularly interesting since it is the cutoff at which both scope orderings are allowed. The surface structure scope ordering is the distributive reading, where there must be

\[
\begin{align*}
\left[ \frac{q_1}{n_1} \right] & = \left[ \frac{4}{2} \right] = 2 \text{ events each performed by} \\
\left[ \frac{q_2}{n_2} \right] & = \left[ \frac{2}{1} \right] \text{ men.}
\end{align*}
\]

Reversing the scope ordering, which is also allowed by the SCOPE rule, we obtain the collective reading where the four men collectively lifted the pianos (one piano at a time, as per the QR in (24)). Note also the case in (30'). The product of a (2-) = 1(2-) = 2, which would make both readings in (30) possible. As the number of students is increased the indirect reading becomes unavailable.

## Transformational Puzzle Revisited

We have thus far avoided discussing the cases of a QR <m_1,m_2> where m_1=m_2. This QR typically reflects a many-to-many relationship. The transformational puzzle is the failure to recognize that the relative ratio of m_1 to m_2 is crucial. Our notion of a quantificational restriction captures exactly this difference. Consider the following RQs:

\[
\begin{align*}
QR\left((\text{Context}), (\text{MemberOf}, \text{Person, Committee}) \right) & = \langle m_1, m_2 \rangle \\
QR\left((\text{Context}), (\text{Love, Man, Woman}) \right) & = \langle w_1, w_2 \rangle \\
QR\left((\text{Context}), (\text{Speak, Person, Language}) \right) & = \langle s_1, s_2 \rangle
\end{align*}
\]

\[^7\text{Note that these QRs are defined on concepts in a semantic network, and therefore QR(Submit, Student, Paper) might be inherited from QR(Submit, Person, Paper).}\]

which are the relevant QRs to the following examples:

(31) John visited a member of every committee
(32) Every man loves a woman
(33) Ten million people speak every Slavic language

We argue that the reason the passive forms of (31) through (33) gradually imply a different meaning is the fact that while m_1 > m_2 (actually m_1 \approx m_2), using the definition of QR in (9), l_1 \gg l_2 and s_1 \gg\gg s_2 (where '\gg' is taken to mean much larger, etc.) In fact, although these relations are all many-to-many, the relative ratio increases to almost many-to-1, which would be "detected" by our SCOPE rule.

## Quantifier Branching Revisited

The QRs implied by example (3), which was discussed in section 1, can be given as follows:

(3) An author of every novel was mentioned in an article by every critic.

QR((Context), (Write, Author, Novel)) = \langle 1^+, w \rangle
QR((Context), (MentionedIn, Novel, Article)) = \langle many, many \rangle
QR((Context), (MentionedBy, Article, Critic)) = \langle b, 1^+ \rangle

The relations involved can be depicted graphically as shown in figure 1 below.

![Figure 1. Branching quantifiers](image)

Note that the intermediate relation, MentionedIn, forms a nucleus while the pair of relations at the two ends are unidirectional. The first and third QRs defined above suggest that the scope must be tightly enforced by the SCOPE rule. The intermediate relation that glues together the entire sentence however suggests that once the scope ordering of both pairs have been decided, the two pairs are independent. More formally, two pairs of quantified term phrases q_1Xq_2C_1 and p_1Yp_2C_2 branch iff

\[
\text{SCOPE}(R(q_2C_1, p_1C_2)) = \text{SCOPE}(R(p_1C_1, q_2C_2)) \text{ where } R \text{ is the intermediate relation between } C_1 \text{ and } C_2.
\]

The SCOPE rule we described above also captures a more subtle semantic dependency that can be illustrated by the following examples:
While the direct reading of (34) could be allowed, the attachment of the prepositional phrase in (36) changes the situation drastically. Essentially, the SCOPE rule captures the fact that while the reference is to “a politician,” in (36) we have “several” (“one” politician from every country), which is the case in (35). According to the SCOPE rule however, the scope ordering in (35) must be reversed.

Concluding Remarks

In this paper we have described a cognitively plausible approach to the resolution of quantifier scope ambiguities. We suggested an alternative to the combinatorial approach where syntactically-motivated rules are used to eliminate possible scope orderings. Instead, we suggest that a most “plausible” reading is first selected by a commonsense reasoning process that uses a dynamically computed value which we termed the quantification restriction. In addition to quantifier scope ambiguity, our model seems to explain some important phenomena that have yet been unresolved.

Two important aspects of this work were not discussed here for lack of space. First, we are currently investigating the classification of general relational schemas that seem to obey the same QR rules. The work of (Lakoff 1987) seems to be relevant in this regard. We are also investigating generic quantification and quantifying over large and potentially infinite domains. Earlier investigations in this regard, which were discussed in (Saba and Corriveau 1995) suggest that the process must be explained in a time and memory constrained model. Further investigation has also suggested that the model must use some form reasoning under uncertainty (Wang 1994).

Finally, we have to acknowledge two important remarks that were made by the reviewers of this paper. First, QR as defined in this paper is a form of a constraint. As such, the SCOPE rule could be seen as a constraint satisfaction rule. Therefore the possibility exists of either under- or over-constraining by the SCOPE rule. Currently we are investigating the ramifications of this on quantifier scope. Finally, the set of quantifiers that we considered in this paper did not include all generalized quantifiers (Barwise and Cooper 1981), such as “many but less than ten.” This extension however has been considered and we are currently extending our formulation of the rule in (23).

Acknowledgements

The authors would like thank three anonymous reviewers for their valuable feedback on an earlier draft of this paper.

References