Towards a Cognitively Plausible Model for Quantification

Walid S. Saba

AT&T Bell Laboratories
480 Red Hill Rd., Middletown, NJ 07748 USA
and
Carelton University, School of Computer Science
Ottawa, Ontario, K1S-5B6 CANADA
walid@eagle.hr.att.com

Abstract

The purpose of this paper is to suggest that quantifiers in natural languages do not have a fixed truth functional meaning as has long been held in logical semantics. Instead we suggest that quantifiers can best be modeled as complex inference procedures that are highly dynamic and sensitive to the linguistic context, as well as time and memory constraints[1].

1 Introduction

Virtually all computational models of quantification are based one some variation of the theory of generalized quantifiers (Barwise and Perry, 1981), and Montague's (1974) (henceforth, PTQ).

Using the tools of intensional logic and possible-worlds semantics, PTQ models were able to cope with certain context-sensitive aspects of natural language by devising interpretation relative to a context, where the context was taken to be an "index" denoting a possible-world and a point in time. In this framework, the intension (meaning) of an expression is taken to be a function from contexts to extensions (denotations).

In what later became known as "indexical semantics", Kaplan (1979) suggested adding other coordinates defining a speaker, a listener, a location, etc. As such, an utterance such as "I called you yesterday" expressed a different content whenever the speaker, the listener, or the time of the utterance changed.

While model-theoretic semantics were able to cope with certain context-sensitive aspects of natural language, the intensions (meanings) of quantifiers, however, as well as other functional words, such as sentential connectives, are taken to be constant. That is, such words *have the same meaning regardless of the context* (Forbes, 1989). In such a framework, all natural language quantifiers have their meaning grounded in terms of two logical operators: $\forall$ (for all), and $\exists$ (there exists). Consequently, all natural language quantifiers

are, indirectly, modeled by two logical connectives: negation and either conjunction or disjunction. In such an oversimplified model, quantifier ambiguity has often been translated to scoping ambiguity, and elaborate models were developed to remedy the problem, by semanticists (Cooper, 1983; Le Pore et al, 1983; Partee, 1984) as well as computational linguists (Harper, 1992; Alshawi, 1990; Pereira, 1990; Moran, 1988). The problem can be illustrated by the following examples:

(1a) *Every student in CS404 received a grade.*
(1b) *Every student in CS404 received a course outline.*

The syntactic structures of (1a) and (1b) are identical, and thus according to Montague's PTQ would have the same translation. Hence, the translation of (1b) would incorrectly state that students in CS404 received different course outlines. Instead, the desired reading is one in which "a" has a wider scope than "every" stating that there is a single course outline for the course CS404, an outline that all students received. Clearly, such resolution depends on general knowledge of the domain: typically students in the same class receive the same course outline, but different grades. Due to the compositionality requirement, PTQ models can not cope with such inferences. Consequently a number of syntactically motivated rules that suggest an *ad hoc* semantic ordering between functional words are typically suggested. See, for example, (Moran, 1988)[2].

What we suggest, instead, is that quantifiers in natural language be treated as ambiguous words whose meaning is dependent on the linguistic context, as well as time and memory constraints.

2 Disambiguation of Quantifiers

Disambiguation of quantifiers, in our opinion, falls under the general problem of "lexical disambiguation", which is essentially an inferencing problem (Corriveau, 1995).

---

[2] In recent years a number of suggestions have been made, such as discourse representation theory (DRT) (Kamp, 1981), and the use of what Cooper (1995) calls the "background situation". However, in both approaches the available context is still "syntactic" in nature, and no suggestion is made on how relevant background knowledge can be made available for use in a model-theoretic model.

Briefly, the disambiguation of "a" in (1a) and (1b) is determined in an interactive manner by considering all possible inferences between the underlying concepts. What we suggest is that the inferencing involved in the disambiguation of "a" in (1a) proceeds as follows:

1. A path from grade and student, s, in addition to disambiguating grade, determines that grade, g, is a feature of student.
2. Having established this relationship between *students* and *grades*, we assume the fact this relationship is many-to-many is known.
3. "a grade" now refers to "a student grade", and thus there is "a grade" for "every student".

What is important to note here is that, by discovering that grade is a feature of student, we essentially determined that "grade" is a (skolem) function of "student", which is the effect of having "a" fall under the scope of "every". However, in contrast to syntactic approaches that rely on devising ad hoc rules, such a relation is discovered here by performing inferences using the properties that hold between the underlying concepts, resulting in a truly context-sensitive account of scope ambiguities. The inferencing involved in the disambiguation of "a" in (1b), proceeds as follows:

1. A path from course and outline disambiguates outline, and determines outline to be a feature of course.
2. The relationship between course and outline is determined to be a one-to-one relationship.
3. A path from course to CS404 determines that CS404 is a course.
4. Since there is one course, namely CS404, "a course outline" refers to "the" course outline.

3 Time and Memory Constraints

In addition to the lingusitic context, we claim that the meaning of quantifiers is also dependent on time and memory constraints. For example, consider

(2a) *Cubans prefer rum over vodka.*
(2b) *Students in CS404 work in groups.*

Our intuitive reading of (2a) suggests that we have an implicit "most", while in (2b) we have an implicit "all". We argue that such inferences are dependent on time constraints and constraints on working memory. For example, since the set of students in CS404 is a much smaller set than the set of "Cubans", it is conceivable that we are able to perform an exhaustive search over the set of all students in CS404 to verify the proposition in (2b) within some time and memory constraints. In (2a), however, we are most likely performing a "generalization" based on few examples that are currently activated in short-term memory (STM). Our suggestion of the role of time and memory constraints is based on our view of properties and their negation We suggest that there are three ways to conceive of

properties and their negation, as shown in Figure 1 below.

P	¬P
1	0
0	1
-	-

(a)

P	¬P
1	0
0	1
-	1

(b)

P	¬P
1	0
0	1
-	0

(c)

Figure 1. Three models of negation.

In (a), we take the view that if we have no information regarding P(x), then, we cannot decide on ¬P(x). In (b), we take the view that if P can not be confirmed of some entity x, then P(x) is assumed to be false[3]. In (c), however, we take the view that if there is no evidence to negate P(x), then assume P(x). Note that model (c) essentially allows one to "generalize", given no evidence to the contrary - or, given an overwhelming positive evidence. Of course, formally speaking, we are interested in defining the exact circumstances under which models (a) through (c) might be appropriate. We believe that the three models are used, depending on the context, time, and memory constraints. In model (c), we believe the truth (or falsity) of a certain property P(x) is a function of the following:

$np(P,x)$ *number of positive instances satisfying P(x)*
$nn(P,x)$ *number of negative instances satisfying P(x)*
$cf(P,x)$ *the degree to which P is "generally" believed of x.*

It is assumed here that cf is a value $v \in \{\perp\} \cup [0,1]$. That is, a value that is either undefined, or a real value between 0 and 1. We also suggest that this value is constantly modified (re-enforced) through a feedback mechanism, as more examples are experienced[4].

4 Role of Cognitive Constraints

The basic problem is one of interpreting statements of the form *every C P* (the set-theoretic counterpart of the wff $\forall x(C(x) \rightarrow P(x))$, where C has an indeterminate cardinality. Verifying *every C P* is depicted graphically in Figure 2. It is assumed that the property P is generally attributed to members of the concept C with certainty $cf(C,P)$, where $cf(C,P)=0$ represents the fact that P is *not* generally assumed of objects in C. On the other hand, a value of cf near 1, represents a strong bias towards believing P of C at face value. In the former case, the processing will depend little, if at all, on our general belief, but more on the actual instances. In the latter case, and especially when faced with time and memory constraints, more weight might be given to prior stereotyped knowledge that we might have accumulated. More precisely:

---

[3] This is the *Closed World Assumption.*

[4] This is similar to the dynamic reasoning process suggested by Wang (1995).

1. An attempt at an exhaustive verification of all the elements in the set C is first made (this is the default meaning of "every").

2. If time and memory capacity allow the processing of all the elements in C, then the result is "true" if np = |C| (that is, if every C P), and "false" otherwise.

3. If time and/or memory constraints do not allow an exhaustive verification, then we will attempt making a decision based on the evidence at hand, where the evidence is based on cf, nn, np (a suggested function is given below).

4. In 3, cf is computed from C elements that are currently active in short-term memory (if any), otherwise cf is the current value associated with C the KB.

5. The result is used to update our certainty factor, cf, based on the current evidence[5].
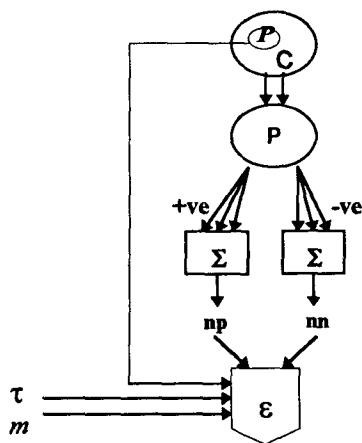

Figure 2. Quantification with time and memory constraints.

In the case of 3, the final output is determined as a function F, that could be defined as follows:

(13) $F_{(C,P)}(nn,np,\varepsilon,cf,\omega) = (np > \varepsilon.nn) \wedge (cf(C,P) >= \omega)$

where ε and ω are quantifier-specific parameters. In the case of "every", the function in (13) states that, in the absence of time and memory resources to process *every* C P exhaustively, the result of the process is "true" if there is an overwhelming positive evidence (high value for ε), and if the there is some prior stereotyped belief supporting this inference (i.e., if cf > ω > 0). This essentially amounts to processing *every C P* as *most C P* (example (2a)).

If "most" was the quantifier we started with, then the function in (13) and the above procedure can be applied, although smaller values for ε and ω will be assigned. At this point it should be noted that the above function is a generalization of the theory of generalized quantifiers, where quantifiers can be interpreted using this function as shown in the table below.

---

[5] The nature of this feedback mechanism is quite involved, and will not be discussed be discussed here.

quantifier	np	nn	ε
every	np = |C|	nn = 0	ε > 0
some	np > 0	nn < |C|	ε > 0
no	np = 0	nn = |C|	ε < 0

We are currently in the process of formalizing our model, and hope to define a context-sensitive model for quantification that is also dependent on time and memory constraints. In addition to the "cognitive plausibility" requirement, we require that the model preserve formal properties that are generally attributed to quantifiers in natural language.

References

Alshawi, H. (1990). Resolving Quasi Logical Forms, *Computational Linguistics*, 6(13), pp. 133-144.

Barwise, J. and Cooper, R. (1981). Generalized Quantifiers and Natural Language, *Linguistics and Philosophy*, 4, pp. 159-219.

Cooper, R. (1995), The Role of Situations in Generalized Quantifiers, In L. Shalom (Ed.), *Handbook of Contemporary Semantic Theory*, Blackwell.

Cooper, R. (1983). *Quantification and Syntactic Theory*, D. Reidel, Dordrecht, Netherlands.

Corriveau, J.-P. (1995). *Time-Constrained Memory*, to appear, Lawrence Erlbaum Associates, NJ.

Forbes, G. (1989). Indexicals, In D. Gabby et al (Eds.), *Handbook of Phil. Logic: IV*, D. Reidel.

Harper, M. P. (1992). Ambiguous Noun Phrases in Logical Form, *Comp. Linguistics*, 18(4), pp. 419-465.

Kamp, H. (1981), A Theory of Truth and Semantic Representation, In Groenendijk, et al (Eds.), *Formal Methods in the Study of Language*, Mathematisch Centrum, Amsterdam.

Kaplan, D. (1979). On the Logic of Demonstratives, *Journal of Philosophical Logic*, 8, pp. 81-98.

Le Pore, E. and Garson, J. (1983). Pronouns and Quantifier-Scope in English, *J. of Phil. Logic*, 12.

Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*. R. Thomason (ed.). Yale University Press.

Moran, D. B. (1988). Quantifier Scoping in the SRI Core Language, In *Proceedings of 26th Annual Meeting of the ACL*, pp. 33-40.

Partee, B. (1984). Quantification, Pronouns, and VP-Anaphora, In J. Groenedijk et al (Eds.), *Truth, Interpretation and Information*, Dordrecht: Foris.

Pereira, F. C. N. and Pollack, M. E. (1991). Incremental Interpretation, *Artificial Intelligence*, 50.

Wang, P. (1994), From Inheritance Relation to Non-Axiomatic Logic, *International Journal of Approximate Reasoning*, (accepted June 1994 - to appear).

Zeevat, H. (1989). A Compositional Approach to Discourse Representation theory, *Linguistics and Philosophy*, 12, pp. 95-131.

325