Semantic Inference in Natural Language: Validating a Tractable Approach

Marc Vilaln
The MITRE Corporation
202 Burlington Rd
Bedford, MA 01730
mbv@hnus mitre org

Abstract

This paper is concerned with an inferential approach to information extraction reporting in particular on the results of an empirical study that was performed to validate the approach The study brings together two lines of research (1) the RHO framework for tractable terminological knowledge representation and (2) the *Alembic* message understanding system There are correspondingly two principal aspects of interest to this work From the knowledge representation perspective the present study serves to validate experimentally a normal form hypothesis that guarantees tractability of inference in the RHO framework From the message processing perspective this study substantiates the utility of limited inference to information extraction

1 Some background

The broad focus of this work has been an attempi to exploit tractable inference in a complex and realistic natural language processing task The task in question is information extraction, that is the process of populating a fixed-field database with information extracted from free-form natural language text The computational framework in which we have explored this research has been the *Alembic* message-understanding system [Aberdeen *et al*, 1993] As with many such systems the extraction process in *Alembic* occurs through pattern matching against the semantic representation of sentences These representations are themselves derived from parsing the input text

That this kind of approach can yield high performance in data extraction is amply documented in [Sundheim, 1992 1993] We have found—as have others—that good results can be obtained with only sketchy sentence semantics (as can happen when there are gaps in the lexicon s semantic assignments) In addition, when the parser normalizes such semantic phenomena as argument passing the number of extraction patterns can be relatively small

Strict semantic pattern-matching is unattractive, however, in cases that presume some degree of inference Consider the following example of what one might term an East-West joint venture (our examples here are either derived or closely inspired from a standard extraction task from the Fifth Message Understanding Conference, that of identifying business partnerships and joint ventures in newswire text)

> [ ] Samsung signed an agreement with Soyuz, the externa]-trade organization of the Soviet Union to swap Korean TV s and VCR s for pig iron [ ]

What makes this sentence an example of the given concept is an accumulation of small inferences that Soyuz is a Soviet entity that signing an agreement designates agreement between the signing parties and that the resulting agreement holds between a Soviet and non-Soviet entity Such examples suggest that it is far preferable to approach the extraction problem through a set of small inferences rather than through some monolithic extraction pattern This notion has been embodied in a number of earlier approaches, e g [Jacobs, 1988] or [Stallard 1986]

The inferential approach we were interested in bringing to bear on this problem is the RHO framework RHO is a terminological classification framework that ultimately descends from KL ONE Unlike most recent such systems, however RHO focuses on terminological inference (rather than subsumption) And whereas most KL ONE descendants sacrifice completeness for computational tractability inference in RHO IS complete in polynomial time if terminological axioms meet a normal form criterion

The primary focus of this paper is thus to show how we applied this idiosyncratic approach to inference to the twin problems of semantic interpretation and data extraction

The second focus of this paper is to present a closely-related empirical study that was actually performed prior to the implementation of inferential data extraction in *Alembic* We undertook this paper study pnor to implementation so as to venfy that the framework could be expected to live up to the data extraction task In particular, we were keen to ensure that the theoretical critenon that guarantees RHO polynomial time completeness was actually met in practice

Giving away the punch line, these findings were encouraging beyond our most optimistic expectations Bringing an implementauon of RHO to bear in a running and externally evaluated version of *AUmbic* further substantiated these findings and we also report briefly on these experiences

Finally the tractability cntenon having survived both the analytic scrutiny of our paper study and the practical scrutiny of implementation, we were led to speculate whether this inferential approach to natural language semantics might somehow be correct at some deep level We conclude the paper with some tantalizing suggestions that this might in fact be precisely the case

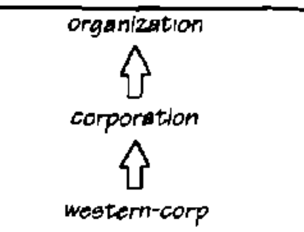

Figure 1 A predicate taxonomy

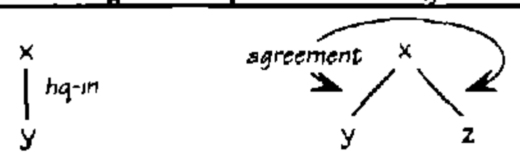

Figure 2 Dependency trees for variables in axioms (1), on the left, and (4), on the right

2 The RHO framework

The RHO framework arose in reaction to the approach to terminological reasoning embodied in most descendants of KL-ONE, e g CLASSIC [Brachman *et al* 1991], BACK [Nebel 1988], LOOM [MacGregor 1991] *etc* This line of work has come to place a major emphasis on computing concept subsumption i e the determination of whether a representational description (a concept) necessarily entails another description In our view, this emphasis is mistaken

Indeed, this emphasis ignores the way in which practical applications have successfully exploited the terminological framework These systems primarily rely on the operation of classification especially instance classification Though subsumption offers a semantic model of classification, it does not necessarily follow that it should provide its computational underpinnings

In addition the emphasis on complete subsumption algorithms has led to restricted languages that are representationally weak Such languages have been the subject of increasingly pessimistic theoretical results from intractability of subsumption [Brachman and Levesque 1984] to undecidability of subsumption [Schmidt-Schauß 1989, Patel-Schneider 1989], to intractability of the fundamental normalization of a terminological KB [Nebel, 1990]

Against this background RHO was targeted to support instance classification and thus departs in significant ways from traditional terminological reasoners The most draconian departure is in separating the normal terminological notion of necessary and sufficient definitions into separate sufficiency axioms and necessity axioms The thrust of the former is to provide the kind of antecedent inference that is the hallmark of classification, e g ,

$$\begin{aligned} western\ corp\ (x) \leftarrow\ & corporation\ (x) \\ & \&\ hq\ in\ (x, y) \\ & \&\ western\text{-}nation\ (y) \end{aligned} \qquad (1)$$

The role of necessity conditions is to provide consequent inference such as that typically associated with inheritance and sort restrictions on predicates e g ,

$$organization\ (x) \leftarrow corporation\ (x) \qquad (2)$$
$$corporation\ (x) \leftarrow western\text{-}corp\ (x) \qquad (3)$$
$$organization\ (x) \leftarrow agreement\ (x, y, z) \qquad (4)$$

Although both classes of axioms are expressed in the same syntactic garb, namely function-free Horn clauses they differ with respect to their inferential import If one thinks of predicates as being organized according to some taxonomy (see Fig 1), then necessity axioms encode inference that proceeds up the hierarchy (i e inheritance) while sufficiency axioms encode inference that proceeds down the hierarchy (i e classification)

The most interesting consequence of RHO s uniform language for necessity and sufficiency is that it facilitates the formulation of a criterion under which classification is guaranteed to be tractable For a knowledge base to be guaranteed tractable the criterion requires that there be a tree shape to the implicit dependencies between the variables in any given axiom in the knowledge base

For the sample axioms above Fig 2 informally illustrates this notion of variable dependencies Axiom (1), for example, mentions two variables, x and y A dependency between these variables is introduced by the predicative term hq in(x,y) the term makes the two variables dependent by virtue of mentioning them as arguments of the same predicate As the axiom mentions no other variables its dependency graph is the simple tree on the left of Fig 1 Similarly in axiom (4) the agreement predicate makes both y and z dependent on x also yielding a tree Finally axioms (2) and (3) lead to degenerate trees containing only x Since all the dependency relations between these variables are tree-shaped the knowledge base formed out of their respective axioms is tractable under the criterion A formal proof that tractability follows from the criterion is in [Vilain 1991], an improved version appears in appendix A below

3 Inferential data extraction

In *Alembic* terminological inference is applied relatively early in the process of semantic interpretation Specifically inference is allowed to take place almost directly on the semantic structures that are produced by the parser-interpreter As our inference axioms are propositional in nature but the semantic representations produced by *Alembic* are not strictly propositional this procedure is mediated by a "propositionalization" phase that maps from the language of interpretations to that of propositions

3 1 Semantic representation in *Alembic*

Alembic produces semantic representations at the popular interpretation level [Alshawi and Van Eijck 1989 Hobbs and Shieber 1987] That is instead of generating fully scoped and disambiguated logical forms, *Alembic* produces representations that are ambiguous with respect to quantifier scoping For example the noun phrase a gold-based ruble maps into something akin to the following interpretation

```
[ [head ruble]
  [quant exists]
  [args NIL]
  [proxy P117]
  [mods { [ [head basis-of]
            [args { P117 [ [head gold]
                           [quant kind]] }]]}]]
```

Semantic heads of phrases are mapped to the *head* slot of the interpretation, arguments are mapped to the *args* slot modifiers to the *mods* slot, and generalized quantifiers to the *quant* slot The *proxy* slot contains a unique variable designating the individuals that satisfy the interpretation If this interpretation were to be fully mapped to a sorted first-order logical form, it would result in the following sentence where gold is treated as a kind individual

∃ P117 ruble basis-of(P117 gold)

Details of this framework are in [Bayer and Vilain, 1991]

3.2 Conversion to propositional form

The propositionalization procedure crucially exploits the proxy variables around which interpretations are built In brief the propositionalization mapping hyper-Skolemizes these proxy variables and then recursively flattens the interpretation's modifiers

For example, the interpretation for a gold-based ruble is mapped to the following propositions

ruble(P117)
basis-of(P117 gold)

The interpretation has been flattened by pulling its modifier to the same level as the head proposition (yielding an implicit overall conjunction) In addition the proxy variable has been interpreted as a Skolem constant, in this case the "gensymed" individual P117 [1] This yields a database of propositions over which inference can be allowed to proceed Say for the sake of argument that we had the following trivial rule

currency(x) ← ruble (x)

Allowing this rule to apply to the propositional form of the interpretation above would yield the conclusion

currency(P117)

3.3 Issues of quantifier scoping and model theory

Note that the interpretation of proxies as Skolem constants is actually hyper-Skolemization, because we perform it on universally quantified proxies as well as on existentially quantified ones Ignoring issues of negation and disjunction, this unorthodox Skolemization process has a disarming model-theoretic justification For a given interpretation with proxy variables v_1 v_n, we simply read v_1 v_n as directly designating some individuals ι_1 ι_n that would satisfy the interpretation in some model μ (wherein the interpretation will have received some unambiguous scoping of its quantifiers) Consider now $P_1(v_1 \; v_n)$ $P_m(v_1 \; v_n)$ the propositions that are mentioned in the interpretation By definition then, $\pi_1(\iota_1 \; \iota_n)$ $\pi_m(\iota_1 \; \iota_n)$ will be satisfied in any such model μ, where the π_i are the interpretations in μ of the P_i The crux is to note that any material implication that is valid in some model of the interpretation will necessarily be valid in all models of the interpretation Since our terminological axioms just perform a simple kind of material implication, it follows that the inferences that they draw will be valid in any model of the interpretation More importantly they will remain so under any scoping of the interpretation's quantifiers

[1] This glosses over event reference which we address in a partly Davidsonian framework as in [Hobbs 1985]

To see this, consider the notorious example 'every man loves a woman " This sentence has two readings, depending on the scoping of the quantifiers the common ∀- reading (every man has a corresponding woman) and the infamous ∃-∀ scoping (there is but one object of affection—Margaret Thatcher or Marilyn Monroe are the usual candidates)

Regardless of the scoping, though, the interpretation of the sentence is propositionalized as

man(P118)
woman(P119)
loves(P118 P119)

Given our reading of proxies, note that under either quantifier scoping P118 will necessarily designate a man P119 will necessarily designate a woman and the loves relationship will necessarily hold between them Now, say we had the following inference rule

romance(x,y) ← loves (x, y) & man (x) & woman(y)

Applying this rule to the propositionalization yields

romance(P118 P119)

Once again this inference is valid regardless of the ultimate scoping selected for the quantifiers This demonstrates a very practical property of our approach namely that it enables inference to be performed over ambiguously scoped text without requiring heuristic resolution of the scope ambiguity (and without expensive theorem proving)

4 Validating RHO

This approach to semantic inference is technically appealing for the simplicity of its inferential framework and for the fact that it can apply so early in the semantic interpretation process Neither characteristic is typical of traditional natural language systems that support inference

Nevertheless, the practical import of our approach would be greatly diminished if it turned out to be (1) too simple to represent useful forms of inference or (2) too computationally onerous in practice These are both empirical questions and as noted above, we strove to address them by first undertaking a paper study in which we applied the approach to a data extraction task It was not particularly obvious how to address the first of these concerns in a clearly quantifiable way, so we were mostly concerned with addressing the issue of computational cost Our goal in particular was to demonstrate that die tractability en tenon we outlined above could in fact be met in practice

Towards this end, my colleagues and I assembled a set of unbiased texts on Soviet economics The validation task then consisted of deriving a set of terminological rules that would allow RHO to perform the inferential pattern matching necessary to extract from these texts all instances of a pre-determined class of target concepts The hypothesis that RHO's tractability criterion can be met in practice would thus be considered validated just in case this set of inference rules was tractable under the en tenon

4 1 Some assumptions

At the ume that we undertook the study, however, the *Alembic* implementation was still in its infancy We thus had to make a number of assumptions about what could be expected out of *Alembic* s parsing and semantic composition components In so doing we took great pain not to require superhuman performance on the part of the parser, and restricted our expected syntactic coverage to phenomena that we felt were well within the state of die art, and that subsequendy were implemented in the running system

In particular, we did not require spanning parses of a sentence As with similar systems *Alembic* uses a fragment parser that produces partial syntactic analyses when its grammar fails to derive S In addition, we exploited *Alembic* s hierarchy of syntactic categories and postulated a number of relatively fine-grained categories thai were not currently in the system This allowed us for example to assume we could obtain the intended parse of "Irish Soviet airline on the basis of the pre-modifiers being both adjectives of geographic ongin (and hence co-ordinable)

We also exploited the fact that the *Alembic* grammar is highly lexicahzed (being based on the combinatorial calegonal framework) This allowed us to postulate some fauly detailed subcategonzation frames for verbs and their nominalizations As is currently the case with our system, we assumed that verbs and their nominalizations are canomcalized to identical semantic representations We also assumed basic competence at argument parsing, a characteristic already in place in the system

4 2 The validation corpus

With these assumptions in mind we assembled a corpus of data extraction inference problems in the area of Soviet *economics The* corpus *consisted of text* passages that had been previously identified for an evaluation of information retrieval techniques in this subject area The texts were drawn from over 6200 Wall Street Journal articles from 1989 that were released through the ACL DCi These articles were filtered (by extensive use of GREP) to a subset of 100-odd articles mentioning the then-exlant Soviet Union These articles were read in detail lo locate all passages on a set of three pre-determined economic topics

- East-West joint ventures these being any business arrangements between Soviet and non-Soviet agents
- Hard currency being any discussion of attempts lo introduce *a* convertible unit of monetary value in die former USSR
- Private cooperatives i e employee-owned enter pnses within the USSR

We found 85 such passages in 74 separate articles (l *2%* of the initial set of articles under consideration)

Among these 47 passages were eliminated as they were just textual mentions of the target concepts (e g the string "joint venture') or of some simple variant These passages could easily be identified by Boolean keyword search—not a particularly insightful validation of a complex NL-based process' Unfortunately this removed all instances of private cooperatives from the corpus, because in these texts, the word ' cooperative" is a perfect predictor of the concept An additional four passages were also removed dunng a cross-rater reliability verification These were all amplifications of an earlier instance of one of the target concepts

target	*occurrences* n	*sufficiency rules r*	*rule density r/n*
Joint *venture*	12	17	14
hard curr	22	13	59

Tablel Summary of experimental findings

eg "US and Soviet officials hailed *the joint project* These passages were eliminated because the corpus collectors had differing intuitions as to whether they were sufficient indications in and of themselves of the target concepts or were somehow pragmatically parasitic upon earlier instances of the target concept The remaining 34 passages required some degree of terminological inference, and formed the corpus for this study

5 Findings

We then set about writing a collection of terminological axioms to handle this corpus We honestly expected that the resulting axioms would not all meet the tractability criterion Natural language is notoriously complex, and even such classic simple KL ONE concepts as Brachman s arch [Brachman and Schmolze, 19851 do not meet the criterion

What we found took us by surprise We came across many examples that were challenging at various levels complex syntactic phenomena, nightmares of reference resolution, and the ilk However, once the corpus passages were mapped to their corresponding interpretations the terminological axioms necessary to perform data extraction from these interpretations all met the criterion

Table I above, summarizes these findings To cover our corpus of 14 passages, we required between two and three dozen sufficiency rules, depending upon how one encoded certain economic concepts and depending on what assumptions one made about argument-passing in syntax We settled on a working set of thirty such rules

Note that this inventory does not include any necessity rules We ignored necessity rules for die present purposes in pan because they only encode inheritance relationships The size of their inventory thus only reflects the degree to which one chooses to model intermediate levels of the domain hierarchy For this study we could arguably have used none In addition necessity rules are guaranteed Lo meet the tractability criterion and were consequently of only secondary interest to our present objectives

Because this validation corpus turned out lo be fairly small me conclusions to be drawn from it could only be considered somewhat preliminary Nevertheless we were very much encouraged by the relative ease with which we were able to write rules that met the tractability cnienon We just didn t seem to find any counterexamples in this data extraction task The fact that the task was both non-trivial and independently motivated was particularly encouraging

This approach to semantic inference and data extraction was also applied in the version *of Alembic* that we fielded for the Fifth Message Understanding Conference As noted the domain in this case was a far more complex model of joint ventures requiring identification of all companies involved in some joint activity their corporate officers (if

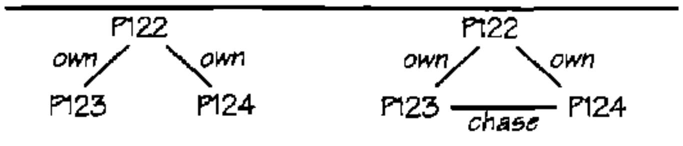

Figure 3 Proxy chains

mentioned) the geographic locations of the companies, the product of the activity, and more—much more

Once again, we found to our pleasure that the inference rules required to cover this expanded task all met the tractability criterion In this particular case, we ended up with a set of 97 rules with coverage breaking down as follows

- Linguistic phenomena (21 rules), covering collocations argument passing and certain contractions
- General knowledge and inference (19 rules) for instance geography or time
- Domain specific inference (56 rules) covering the particulars of the domain

In both our paper study and our MUC 5 system the rules we ended up writing were largely pedestrian even bonng Most rules have three terms two unary predicates and a relation linking them Predicate valence is never greater than three and only a handful of axioms yield dependency trees more exotic wan a simple linear sequence of variables

6 Some speculations

Both our paper study and our experience with MUC 5 can be taken as empirical indications that the tractability criterion in RHO is indeed a realistic restriction It is our belief that tractable non trivial inference is thus a practical reality, even in applications as complex as data extraction systems Buoyed as we were by these results we began to question whether they might not be due to some general character istics of language In fact this seems to be tantalizingly so

In particular the intractable class of axioms is closely implicated with anaphora resolution one of the classic hard problems in natural language processing To be specific, axioms that violate the tractability en tenon can only be satisfied by sentences that display some kind of anaphora such as pronouns or definite references

This can be seen by considering die way in which chains of proxy individuals are formed in the process of semantic composition In particular, proxies are introduced by the heads of noun phrases, and are chained together eitfier by application of the verb phrase or by noun phrase modifiers For example the sentence ' a man owns a cat" introduces two proxies mat are chained by the verb own '

man(P120)
cat(P121)
own(P120 P121)

The same propositionalization is produced with the analogous NP/relativc clause combination "a man who owns a cat ' Similar chains are also produced for prepositional phrase attachment participial vp modifiers and other forms of NP modification Note however, that in the absence of anaphora, these chains are constructed independently No chain may refer to variables drawn from another chain an observation, noted in another context by Haddock [1992]

Consider for example 'a man who owns a cat and who owns a dog which propositionalizes as

man(P122)
cat(P123)
own(P122 P123)
dog(P124)
own(P122 P124)

The chaining between these variables is a simple tree (see Fig 3, left tree) For the chaining to yield a true graph would in mis case require the dog in one modifier to refer to the cat in the other as would happen if the KB contained

chase(P124 P123)

Adding this proposition to the KB yields a circular chaining of proxies, as in the nght tree of Fig 3 Crucially such circularities can only arise through anaphoric reference as in ' a man who owns a cat and owns a dog thai chases the cat, or like cases (' chases it/his cat/that cat)

This only addresses the construction of circular proxy chains among the propositionalizations of the linguistic input not among the terms of inference axioms The crucial observation however is that absent such circular proxy chains in the KB axioms that are outside the criterion fail to be satisfied Indeed an axiom is outside the criterion jnst in case the variables in its terms exhibit non-tree-like dependencies as in die following silly rule

hapless(x) ← own(x z) & own(x, y) & chase(y z)

For these terms to match against the linguistic KB propositions in the KB must exhibit corresponding circularities which only happens if the linguistic input is anaphoric

It is especially important not to assume that the converse of this result holds That is just because axioms that fail the criterion can only be satisfied by the propositionalizauon of anaphora it is not the case that anaphoric use of language leads to intractability Cntenon-passing axioms lead to tractable inference regardless of whether the facts in the prepositional KB were derived from surface anaphora

Interestingly, the very concepts encoded by criterion-failing axioms are themselves of a complex flavor Indeed, attempting to paraphrase such axioms in English in turn requires anaphora The silly rule, for example comes out as a hapless (person) is one who owns a (presumed) pet and owns another (presumed) pet that chases *the first pet/it/the first one/that first pet/* " One can simply not render this rule in English without resorting to pronouns or the ilk

It is truly tantalizing that the cases where terminological inference in RHO is computationally hard align with such linguistically hard phenomena as anaphora Perhaps this alignment may help explain the dearth of intractable terminological axioms in our paper study and in our MUC 5 system The alignment also suggests that Brachman may have been more nght than he thought in the early days of KL-ONE, when he suggested that terminological reasoning was really about the semantics of noun phrases

Such alignments are also fertile ground for wild speculation about the nature of language or reasoning Yielding to cool-headed restraint, though one may still conclude that useful inference in natural language is less intractable man was previously assumed And one is no less justified in echoing Alice's ineffable words, ""Cunouser and cunouser'

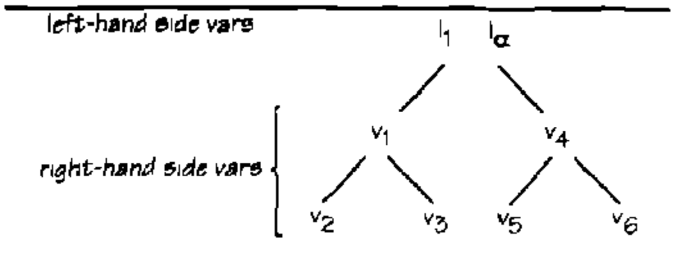

Figure 4 A dependency graph

Acknowledgments

Much gratitude is owed John Aberdeen for tireless perusal of the Wall Street Journal Thanks also to Remko Scha Bill Woods Steve Minton Dennis Connolly and John Burger

Appendix A Proof of tractability

To demonstrate the validity of the tractability criterion we only need consider the computational cost of finding all instantiations of the right hand side of an axiom In general finding a single such instantiation is NP complete, by reduction to the conjunctive Boolean query problem [Garey and Johnson 1979] Intuitively this is because general function-free Horn clauses can have arbitrary interactions between the variables on the right-hand side i e their dependency graphs are fully cross-connected, as in

$$R(v_1\, v_2)\ \&\ R(v_1\, v_3)\ \&\ R(v_2\, v_3)\ \&\ R(v_1\, v_4)\ \&\ R(v_2\, v_4)$$

Intuitively again, verifying the instantiation of a given variable in a rule may require (in the worst case) checking all instantiations of all other variables in the rule Under the usual assumptions of NP-completeness no known algorithm exists that performs better in the worst case than enumerating all these instantiations As each variable may take on as many as κ instantiations where κ is the number of constants present in the knowledge base the overall cost of finding a single globally consistent instantiation is $O(\kappa^\xi)$ where ξ is the number of variables in the rule The resulting complexity is thus exponential in ξ which itself varies in the worst case with the length of the rule

Consider now an axiom that satisfies the tractability criterion yielding a graph such as that in Fig 4 By definition the root of the graph corresponds to all the variables on the left-hand side and all other nodes correspond to some variable introduced on the right-hand side The cost of finding all the instantiations of the root variables is bounded by κ^α where α is the maximal predicate valence for all the predicates appearing in the database The cost of instantiating each non-root variable v is in turn bounded by $\alpha\kappa^\alpha$, corresponding to the cost of enumerating all possible instantiations of any predicate relating v to its single parent in the graph

The topological restriction of the criterion leads directly to the fact that the exponent of these terms is a low-magnitude constant, α, rather than a parameter ξ, that can be allowed to grow arbitrarily with the complexity of inference rules The topological restriction also leads to the fact that these terms contribute *additively* to the overall cost of finding all instantiations of a rule This overall cost is thus bounded by $\kappa^\alpha + \underbrace{\alpha\kappa^\alpha + \quad + \alpha\kappa^\alpha}_{\xi}$, or $O(\xi\alpha\kappa^\alpha)$

Finally, note that with the appropriate indexing scheme *finding all consequents of all rules only adds a multiplicative* cost of ρ, where ρ is the total number of rules, yielding a final overall cost of $O(\rho\xi\alpha\kappa^\alpha)$ It is often assumed that predicates in natural languages have no more than three arguments so this formula approximately reduces to $O(\kappa^3)$

References

Aberdeen J Burger J Connolly D Roberts S and Vilain, M (1993) MITRE-Bedford Description of the Alembic system as used for MUC 5 In [Sundheim 1993]

Alshawi H and Van Eijck, J (1989) 'Logical forms in the core language engine In *Prcdgs of ACL89* Vancouver BC

Bayer S and Vilain M (1991) 'The relation-based knowledge representation of King Kong *Sigart Bulletin* 2(3)

Brachman, R J Borgida A McGuiness D L & Patel-Schneider, P F (1991) 'Living with CLASSIC' In Sowa J *ed Principles of Semantic Networks* San Mateo CA Morgan Kaufmann

Brachman R J and Levesque H (1984) The tractability of subsumption in frame-based description languages In *Prcdgs of AAAI84* Austin TX

Brachman R J and Schmolze J (1985) An overview of the KL ONE knowledge representation system *Cog Sci* 9

Garey, M R and Johnson D S (1979) *Computers and Intractability* New York W H Freeman

Haddock, N J , (1992) Semantic evaluation as constraint network consistency In *Prcdgs of AAAI92* San Jose CA

Hobbs, J R (1985) '*Ontological promiscuity* In *Prcdgs of ACL85* Chicago IL 119–124

Hobbs J R and Shieber S M (1987) 'An algorithm for generating quantifier scopings *Comput Linguistics* 13(1 2)

Jacobs P S (1988) 'Concretion Assumption-based understanding In *Prcdgs of COLING88* Budapest

MacGregor R (1991) 'Inside the LOOM description classifier *Sigart Bulletin* 2(3)

Nebel B (1988) Computational complexity of terminological reasoning in BACK' *Artificial Intelligence* 34(3)

Nebel B (1990) Terminological reasoning in inherently intractable *Artificial Intelligence* 43

Patel-Schneider P F (1989) Undecidability of subsumption in NIKL *Artificial Intelligence* 39

Schmidt-Schauß M (1989) Subsumption in KL ONE is undecidable In *Prcdgs of KR89* Toronto ON

Stallard D G (1986) A terminological simplification transformation for natural language question answering systems In *Prcdgs of ACL86* New York NY

Sundheim B *ed* (1992) *Prcdgs of the Fourth Message Understanding Conf* (MUC-4) McLean VA

Sundheim, B *ed* (1993) *Prcdgs of the Fifth Message Understanding Conf* (MUC 5), Baltimore MD

Vilain M (1991) Deduction as parsing tractable classification in the KL ONE framework In *Prcdgs of AAAI91* Anaheim CA