Towards Adequate Knowledge and Natural Inference

Lenhart Schubert and **Jonathan Gordon** and **Karl Stratos** and **Adina Rubinoff**
Department of Computer Science
University of Rochester
Rochester, NY, USA
{schubert, jgordon, jlee164, arubinoff}@cs.rochester.edu

Abstract

We are striving to create broad bases of inference-capable lexical knowledge and world knowledge, and at the same time are experimenting with "natural" inference using a reasoning engine (EPILOG) that employs an event-oriented, natural-language-like representation, Episodic Logic (EL). The goal is to be able to support domain-independent language understanding and commonsense inference. At this point, we have several hundred core lexical axioms, many millions of of generic "factoids" of varying quality derived by superficial interpretation of text corpora, 1.5 million quantified formulas obtained by "sharpening" such factoids, and thousands of conditionals derived from discourse cues. We are able to demonstrate various inferences, including ones that are the focus of recent work in "natural logic" (NLog), but also many that are beyond the scope of NLog. However, major gaps remain in bridging from language to deep understanding and inference.

1 Introduction

Human-level intelligence will require knowledge representations and inference methods as rich and subtle as those employed by humans. A long-standing goal of our group at the University of Rochester has been to develop knowledge representations and inference methods that would be adequate for broad-spectrum language understanding and commonsense reasoning. The outcome of over two decades of sporadic work on this goal has led to a general natural language-like (and Montague-like) knowledge representation, Episodic Logic (EL), and two versions of inference engines for this representation, EPILOG 1 and EPILOG 2. This work is still continuing, but over the last decade, we have also devoted much of our collective effort to knowledge accumulation, in order to alleviate the infamous knowledge acquisition bottleneck. We have developed several new ways of accumulating quantified knowledge, and have recently demonstrated on a modest scale that at least some of this knowledge can be used effectively for drawing commonsense conclusions from given facts. Some of the reasoning emulates Natural Logic (NLog) principles, but it also allows for integration of lexical knowledge and world knowledge.

We will begin with a brief review of the EL representation and of EPILOG 1 and 2. In sections 3 and 4 we then describe our work to date on acquiring various sorts of lexical knowledge and world knowledge respectively. In section 5, we report on some recently instantiated types of inference in EPILOG, and in section 6 we sum oup the status of our work, and comment on related work and on remaining challenges confronting deep understanding by machines.

2 Episodic Logic (EL), EPILOG, & NLog

Figure 1 provides a glimpse of both the EL representation and the overall EPILOG 1 architecture (Schubert and Hwang 2000). Note that the representation of the sentence, "A car crashed into a tree" is quite language-like, in its use of restricted quantifiers and NL-like predicate-argument structure. (Predicates are infixed, i.e., follow the subject argument, for readability.) A difference from NL lies in the use of explicit episodic terms standing for events of situations, where these are associated with sentences that describe them. For example, note the variable 'e' in the sample formula, standing for the event of the car x crashing into tree y, and connected to its characterizing sentence by the characterization operator '**'. This variable enables anaphoric reference to the event in question, and can participate in temporal, locative, causal or other relations. If the indicated sentential formula were supplied to EPILOG, and the knowledge base contained a conditional formula to the effect that if a car crashes into a tree, the driver of the car may be hurt or killed, then the expected conclusion about the driver would be inferred by EPILOG, as indicated.

EPILOG 1 performs both forward (input-driven) and backward (goal-driven) inference, some of it probabilistic, and it is systematically aided by a dozen specialist subsystems for efficient taxonomic, partonomic, temporal, etc., inference. Its capabilities have been demonstrated in a variety of small domains – fairy tale fragments, terrorist incidents, short planning dialogues, and others. However, it suffers from various "blind spots" particularly in backward inference, and its meta-inference facility is awkward; because of its incremental mode of development in parallel with development of EL, the code is complex and sparsely commented, making it hard to amend or extend, and this led to the implementation of EPILOG 2 by Fabrizio Morbini. EPILOG 2 performs goal-directed inference very effectively, and despite its trans-FOL representational abilities, holds its own against state-of-the-art theorem provers for problems stated in FOL. What is still lacking is a full forward inference capability, probabilistic inference, integration with the specialists, and an associative

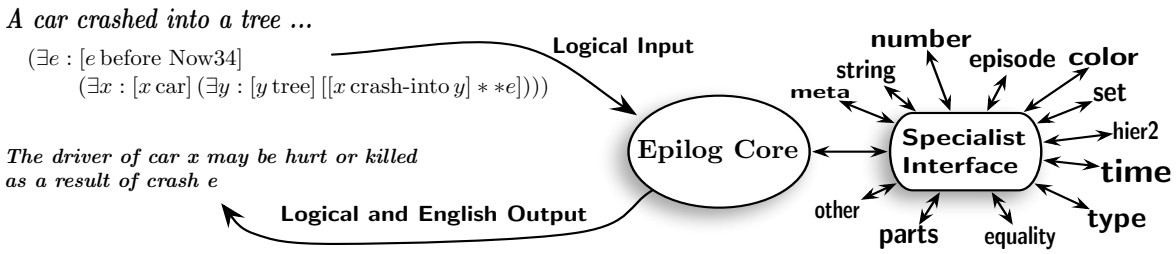

Figure 1: Episodic Logic and the EPILOG system

retrieval mechanism suitable for a KB of tens of millions of knowledge items, as will be needed eventually.

The following are two additional examples of the EL representation, illustrating some of the ways in which it transcends FOL.[1]

- Restricted quantifiers:
  Most laptops are PCs or MACs
  ```
  (Most x: [x laptop] [[x PC] or [x MAC]])
  ```
- Modification and reification:
  "He firmly maintains that aardvarks are nearly extinct"
  ```
  (Some e: [e at-about Now17]
          [[He (firmly (maintain
                        (that [(K (plur aardvark))
                               (nearly extinct)])))] ** e])
  ```

Both examples illustrate restricted quantification (a variant of generalized quantification). The second example is designed to illustrate how the allowance for predicate-modifying operators (such as 'firmly', 'plur', and 'nearly'– the last being intensional) and reifying operators (such as 'that' and the kind-forming operator 'K') keep the syntactic and semantic structure of logical forms very close to those of the sentences they formalize. The logical syntax and semantics have much in common with Montague's intensional logic, apart from the first-order approach to quantification, and the "lowering" of sentence and predicate intensions occurring in predicate subject or object positions to individuals, by means of type-shifting operators like 'that' and 'K'.

As will be seen shortly, EL and EPILOG also allow *substitutional quantification* and *quasi-quotation* (transparent to substitution for syntactic metavariables). This capability has proved crucial not ' only for implementing a degree of self-awareness and "I-would-know-it-if-it-were-true" types of inference (Morbini and Schubert 2008), but also for emulating *Natural Logic* (NLog).

NLog entailments are enabled by lexical meaning postulates. In particular, the idea is to allow replacement of sentential subexpressions by semantically more general (lexically entailed) ones in positive-polarity environments, and by more specific (lexically anti-entailed) ones in negative-polarity environments. For example, we have the entailments

"Several trucks are on their way"
|= "Several large vehicles are on their way"
"If a large vehicle is on its way, turn it back"
|= "If a truck is on its way, turn it back".

In the first sentence, we can replace 'trucks' by the more general term 'large vehicles' because 'several' is upward monotone (upward entailing) in its restrictor, so that 'trucks' lies in a positive environment in that case. In the second sentence we can replace 'large vehicle' by the more specific 'truck', because 'if' is downward monotone in its complement, while the determiner 'a' is upward monotone in its restrictor, so that 'large vehicle' lies in a negative environment. The meaning postulate needed here is simple, not requiring syntactic metavariables:

```
(all x [x truck] [x ((attr large) vehicle)]).
```

(Here 'attr' is a type-shifting operator transforming a monadic predicate into a predicate modifier, one that maps monadic predicates to monadic predicates.) Note incidentally that the above entailments hold regardless of the vagueness of 'several' and the context-dependent meaning of 'on their way', as long as the meanings of these terms are the same on both sides of the entailment. This tolerance of underspecification is shared by EPILOG, though as we note later, genuine understanding and common sense can only tolerate underspecification to a limited extent.

Another important aspect of NLog is its exploitation of *implicatives* and *factives*. In particular an implicative verb such as 'manage (to)' and 'agree (to)' implies that their respective subjects actually did what they managed or agreed to do. As is characteristic of implicatives, these implications are reversed in a negative polarity environment; i.e., if the subjects did not manage/agree to do something, they did not do it (at least at that time). We can formalize these implications in EL in the following way:

```
(all_pred p (all x [[x manage (Ka p)] => [x p]])),
(all_pred p (all x [(not [x manage (Ka p)]) => (not [x p])])).
```

Here 'all_pred' quantifies over all syntactic substitutions that replace 'p' with a monadic predicate. 'Ka' forms a kind of action or attribute from a predicte (it is definable in terms of 'K' and episodic operators). Similar axioms hold for many other implicative verbs (these are typically subject-control verbs with an infinitive complement). In other cases, such as 'forget (to)' and 'refuse (to)', we have a negative implication in a positive environment and a positive implication in a negative environment. For example, if Mary did not refuse to have dessert, she presumably did have dessert (at least as a defeasible conclusion). Other implicative "signatures"

[1] Actually, we append syntactic-category extensions to atoms for disambiguating their types, e.g., 'laptop.n' or 'hurt.v', but omit these here for readability. Also, square brackets become round in EPILOG– they serve here to distinguish infix formulas visually. The colons after quantified variables indicate the presence of a restrictor, but are also omitted in the EPILOG representation.

are possible (e.g., consider 'hesitate (to)'), but we leave the discussion there.

Factive verbs are ones like 'know (that)' and 'realize (that)', taking a *that*-clause as complement, and they differ from implicatives in that their entailments are positive even when the verbs lie in a negative environment. In other words, their entailments are *presuppositions*. In this case we use forward inference rules like

```
(all_wff w (all x ((x know (that w)) ---> w))),
(all_wff w (all x ((not (x know (that w))) ---> w))).
```

There are good reasons for using inference rules here rather than axioms, as the latter would render all formulas provable (Stratos, Schubert, and Gordon 2011). Examples of *antifactives* are 'pretend (that)' and 'lie (that)'. These call for rules similar to those above but with (not w) as the conclusion.

We will provide examples of inferences obtained by EPILOG 2 using axioms and rules like those we have shown, after our discussion of knowledge acquisition. Note finally that implicativity and factivity extend beyond verbs to nouns and adjectives, and to various phrasal constructions; e.g., 'success (in)', 'failure (to)', 'fact (that)', 'pretense (that)', 'sad (to)', 'have the gall (to)', 'it is a pity (that)', and many others.

3 Towards extensive lexical semantic knowledge

We discuss our approaches to acquiring lexical knowledge under three subheadings; in each case, we rely to some extent on existing resources, and to some extent on manual or semi-automatic methods.

Entailment, synonymy, and exclusion relations Many computational semantic researchers turn first to WordNet for lexical semantic knowledge, as its synonym clusters, hypernym hierarchy, antonyms, partonomic relations and other annotations provide a rich source of basic entailments. We too evaluated WordNet 10 years ago as a source of type subsumption (between hypernym-hyponym pairs) and exclusion knowledge (between sister synsets, for non-role nominals), and found it too unreliable at the time (Kaplan and Schubert 2001). However, inspired by the method of classifying pairwise word relations in (MacCartney and Manning 2008), and encouraged by continued upgrading of WordNet, we recently took another look at the feasibility of building a KB of such pairwise relations (Schubert, Van Durme, and Bazrafshan 2010). Like MacCartney and Manning, we used WordNet path features as well as miscellaneous features such as word morphology and distributional similarity to classify the relation between lexical items within distributional similarity clusters (as provided by Patrick Pantel) as synonymy, (proper) entailment, opposition, nonexhaustive exclusion, or unrelated. Accuracy ranged from 65% to 90% for verbs, adjectives, and nominals. Since the WordNet features turned out to be the decisive ones, with other features helping little to boost scores, we interpret the results as reflecting the accuracy of WordNet itself on entailment relations that can be extracted from it. However, considerable work remains to be done to obtain a reliable and comprehensive set of pair-relations. The process is computation intensive, and there are questions such as what to do about concepts with implicit arguments. For example, how do we formalize that a "price" is a "value"?

Implicative and factive verbs We have created a collection of about 250 implicative, factive, and belief- or want-entailing lexical items, drawing on work by Nairn, Condoravdi, and Karttunen (2006) and Danescu-Niculescu-Mizil, Lee, and Ducott (2009). We also expanded the collections using Wordnet synonyms and cognates within VerbNet classes. Using their implication signatures, we generated axioms and inference rules for the lexical items (some of them are multi-word verbs) of the type described earlier. As we describe in section 5, we have successfully tested the resulting knowledge using EPILOG 2.

Event-oriented lexical axioms Most existing verb frame lexicons provide only very weak information about the meaning of eventive verbs, such as semantic role labels or type constraints (animate, phys-object, destination, and the like) on arguments, or indications that the agent or object are in motion in the event, or that an argument may specify a result state. We do not find out from such sources that dressing oneself involves putting on clothing, that picking up an object involves grasping and lifting it, that requesting people to do something conveys to them the requester's desire that they do it, and so on.

Our current efforts to provide such lexical information began with assembly of about 100 verb senses that seem to qualify as "primitives", in the sense that even very small children surely possess the corresponding concepts, and that numerous verbal concepts have entailments involving them. For example, any child understands the concept of grasping or letting go of an object (as well as picking one up or dropping one), or of wanting something, of walking, crying, hurting (someone), or asking someone to do something. At the same time, for example, 'grasp' can be used in the axiomatization of certain senses of 'grip', 'clench', 'seize' (near-synonyms), and we find it among the entailments of 'pick up', 'hug', 'cuddle', 'strangle', 'tackle (someone)', 'catch (a fast-moving graspable object)', and so on. For guidance in identifying plausible primitives, we used not only such intuitions but also the repertoire of about 20 semantic predicates found in VerbNet (e.g., begin, exist, force), and some 65 VerbNet class names (e.g., break, carry, fill, learn, own, pour, and stop). Examples of axioms for primitives are the following:

```
(all x (all y (all e: [[x lift y] ** e]
  [[x ((adv-a upwards) (move-trans y))] * e]))),

(all_pred p (all x (all y: (all e1 [[x ask-of y (Ka p)] ** e1]
  [[x convey-info-to y
     (that [[x want-tbt
              (that (some e2 [e2 right-after e1]
                             [[y p] ** e2]))] @ e1])] * e1]))))),
```

i.e., lifting something entails moving it upward (and this moving event is a part of the lifting event), and asking someone to do something entails conveying to them that one wants them to do it. We currently have about 110 such tentative axioms for presumed primitives. We are also formulating *disambiguation rules*, such as the following ones for catching a physical object (e.g., a ball) or catching a communicable illness (e.g., a cold):

```
[X_anim? catch Y_phys-obj?] -->.9 [X catch-obj Y],
[X_anim? catch Y_communicable-illness?] -->.8 [X catch-illness Y].
```

The idea is that these pattern-based rules can be applied in any embedding context, regardless of polarity or intensionality (unlike axioms).

This approach reflects our view that disambiguation is not based on the fit of the resulting sentence meaning to the world, but simply the fit of its parts to familiar patterns of meaning – which is what our rules try to capture. Variable constraints, attached with an underscore and tagged with a question mark, are intended to be executable predicates that access knowledge about the arguments bound to the variables. The numbers on the arrows can be thought of as probabilities that the choice of verb sense provided by the rules are correct, given only the satisfaction of the antecedent pattern.

We see our work on primitive axiomatization as the first stage of a more comprehensive verb axiomatization, using VerbNet in the following way. We associate a general axiom schema with each VerbNet class, employing primitive predicates as much as possible for their eventive aspects. (Some of the classes already correspond to primitives, but that does not prevent supplying entailments in terms of other primitives and non-eventive predicates.) For example, for the 'create' class we have the schema (including type extensions on predicates for clarity here)

```
(all x (all y: [y A] (all e: [[x VERB y] ** e]
                             [[x make.v y] ** e]))),
```

where A and VERB are parameters to be instantiated separately for particular members of the verb class, and 'make.v' is a primitive understood in the sense of causing something to exist. Examples of pairs of values for these parameters are (coin.v, (K coin.n)) (i.e., to coin a kind of coin is to cause that kind of coin to exist), (compute.v, piece-of-info.n) (i.e., to compute a piece of information is to cause that piece of information to exist), and similarly (concoct.v, mixture.n), (concoct.v, plan.n))), etc. We have treated only a few classes so far, but based on that the development program seems feasible, modulo some adjustments and partitioning of current classes.

Finally under this subheading we mention our approach to handling PP complements and adjuncts of verbs that intuitively modify the verb meaning in a systematic, compositional fashion. (This excludes cases like 'abide by', 'yearn for', believe in', etc., where we should form predicates such as 'abide-by', 'yearn-for', etc.) Rather than separately including such PPs in the axiomatics of the verb, we treat them as modifiers. For example, the instrumental PP in "Mary poked John with a ruler" would initially be applied to the predicate (poked John) as follows (where tense has been neglected, <a ruler> is an unscoped determiner plus restriction predicate, and 'with-instr' is the result of applying disambiguation rules):

```
((adv-a (with-instr <a ruler>)) (poked John)).
```

Here 'adv-a' transforms a predicate (usually a PP predicate) into a predicate modifier applicable to an action predicate (here, (poked John)). We use systematic deindexing and scoping rules in transforming initial logical forms to ones with explicit episodic variables (see, e.g., (Schubert and Hwang 2000)). Here, this process leads to an instrumental predication conjoined with the 'poke'-predication,

```
[(Mary . e) with-instr y],
```

where e is the episodic variable (the poking event), (Mary . e) is the agent-event pair (an action, as distinct from an event), and y is the variable bound by the determiner 'a' (i.e., it refers to the ruler). This is obviously rather similar to what would be obtained with a thematic-role-based approach that treats the instrumental role as part of the verb frame, but the point is precisely that we avoid having to exhaustively specify the roles that may be added by PP complements and adjuncts to a verb.

4 Towards extensive knowledge of semantic patterns and the world

Our efforts to accumulate shallow general "factoids" about the world, as a preliminary to obtaining deeper, more complex world knowledge, began more than a decade ago (Schubert 2002). The idea was that specific sentences such as

> "Kelly bought some cheap food at the local convenience store to tide him over the weekend"

reveal what are likely to be recurrent types of events and situations in the world, such as that

> A person may buy food;
> Food may be cheap; and
> Food may be at a convenience store.

The Knext (KNowledge EXtraction from Text) system that we developed obtains such general factoids by interpreting parsed sentences with compositional semantic rules (about 80 such rules). Each rule matches the right-hand side of some class of phrasal rules with a somewhat enhanced type of regular expression, and forms an EL logical form by composing the logical forms of the phrasal constituents (possibly omitting some constituents). Pieces of the logical form are selected that look promising for forming factoids, some modifiers are dropped, named entities are abstracted to general types using gazetteers and other methods, and determiners are in most cases replaced by a general indefinite. The resulting factoids are automatically rendered into English, in "possibilistic" form (using "may" or "can").

We eventually obtained tens of thousands of factoids from the Brown corpus (Schubert and Tong 2003), millions from the British National Corpus (BNC) (e.g., (Van Durme, Qian, and Schubert 2008)), and hundreds of millions from Wikipedia and weblogs (Gordon, Van Durme, and Schubert 2010). The cited papers describe the character of the factoids, filtering methods, and various comparative evaluations. Overall, roughly four out of five factoids are judged to be reasonable, potentially useful general claims about the world. Their potential utility is two-fold: (1) They should be useful as semantic patterns that could guide a parser towards better choices; and (2) They could be manipulated in various ways to obtain stronger abstractions – ones that could be useful for commonsense inference.

To illustrate the first point, familiar ambiguous sentences like

> "Time flies like an arrow", or
> "John saw the bird [with binoculars / with yellow tail feathers]",

could be parsed reliably if we knew, as a familiar pattern, that "Time may fly" (as indeed is found in the Knext KB), whereas there are no patterns containing a compound nominal 'time flies', nor any about people timing flies. Similarly patterns

of type "A person may see with binuculars", or "a bird may have feathers" should help greatly with the disambiguation of the two variants of the second sentence.

Our plans to use such methods await the development of a parser that is more transparent and more easily guided than current off-the-shelf statistical parsers. Our emphasis in this paper is on the second goal, viz. manipulating factoids to obtain inference-capable knowledge. We briefly report on two such methods, as well as a method of obtaining conditional (if-then) knowledge directly from certain types of parsed sentences containing revealing discourse markers such as 'but' and 'hoping to'.

Abstracting from clusters of factoids Factoids containing the same verb often convey similar ideas, as in the two factoids

A child may write a letter, and
A journalist may write an article.

In (Van Durme, Michalak, and Schubert 2009), the WordNet hypernym hierarchy is used to abstract from clusters of such factoids to ones such as

Generally, if X writes Y, then X is a person or organization; and
Generally, if X writes Y, then Y is a communication.

In other words, these more general, quantified assertions provide selectional preferences in a logically useful form. The abstraction method, for the cluster of nominals for a particular argument position, intuitively consists of searching upward in the sense hierarchy for a superordinate concept, or *small* disjunction of superordinate concepts, dominating *some* sense of *most* of the nominals in the cluster, without dominating too many nominals (especially common ones) not in the cluster. For example, a certain WordNet synset for 'communication' (in writing) turned out to meet this criterion for the object of 'write'. The added benefit of the abstraction is that it selects specific senses of the nominals it abstracts from, in the process of finding hypernym paths to shared abstractions. Thus one of the five senses of 'letter' (namely the type that gets mailed) is selected for that noun as the object of 'write', and similarly one of the four senses of 'article' (the literary kind) is selected for that object.

The evaluations performed for this abstraction method showed the abstractions to be generally reasonable, and also showed the resulting nominal disambiguations to be superior to baseline methods that just choose the most common sense of a word. However, the work remains incomplete in two respects: Because the abstraction process is computationally demanding (as a result of the many combinations of superordinate concepts that need to be searched for large nominal clusters), it has not been applied to a very large set of factoids; and for many verbs, specially light verbs, it would be desirable to abstract subject-object *combinations* of argument types, rather than subject types and object types separately. For example, 'carry' intuitively allows for such type combinations as

A person may carry a hand-portable object,
A person may carry a grudge,
A vehicle may carry persons or freight,
A road may carry traffic,
A newspaper may carry a story,
etc.,

and it is scarcely helpful to know that if something is carried, it is generally a hand-portable object, or a grudge, or persons, or freight, or traffic, or a story, etc. This is a challenging problem, since the categories need to be found jointly, and often are ones specific to the verb, not likely to be found in WordNet. (For example, miscellaneous hand-portable objects such as guns, umbrellas, newspapers, briefcases, shopping bags, and so on, do not fall under any one existing WordNet synset that excludes non-portable objects.)

"Sharpening" factoids The most productive method of creating general quantified formulas we have developed so far is that reported in (Gordon and Schubert 2010). It consists of transforming factoids such as

A tree may have a branch, or
A person may have a sandwich,

into "sharpened" formulas such as the following, using engineered transformation rules:

```
(all-or-most x: [x tree]
             (some y: [y branch]  [x has-as-part y])),
(many x: [x person]
             (at-least-occasional e
                      (some y: [y sandwich] [[x eat y] ** e]))),
```

i.e., All or most trees have some branch as a part, and many people at least occasionally eat a sandwich. Note that this allows the prediction for a given tree that it probably has at least one branch, and for a given person that he or she may well occasionally eat a sandwich.

The rules match input templates to factoids, and produce a sharpened formula via an output template. An example of an input template is

```
((1_det? 2_plant?) have.v (3_det? 4_plant-part?)),
```

and the output template would be much like the first formula above, except for having the match variables '2_' and '4_' in place of 'tree' and 'branch' respectively. The "questioned" type restrictors attached to match variables are functions evaluated with access to various resources such as WordNet, VerbNet, and corpus frequency information. For example WordNet allows classification of 'tree' as a 'plant', and (via partonomic information) classification of 'branch' (in one sense) as a plant part. (Note that we can find part-of relations that are not in WordNet, for example that a contraption may in many cases have a button as a part – where WordNet need not supply 'button' as a part of a contraption, only as part of some artifact such as a shirt, doorbell or cell phone.)

An important distinction that can also be approximately implemented is that between verb senses that express enduring properties or relationships (so-called *individual-level* properties such as having something as a part, or having someone as a particular type of relative) from ones that express transient or telic/ episodic properties (so-called *stage-level* properties such as eating or talking or receiving something). This is the basis for the two very diffeent types of sharpened formulas above.

The rules also make use of VerbNet, for example to identify *nonrepeatable* (or rarely repeatable) actions and events such as being born, being killed, marrying, graduating, and so on. after identifying a few such verbs, we can collect many more from their VerbNet "classmates". Corpus frequencies, as well, help to decide on how frequent a type of event should be posited to occur for particular argument types.

At the time of writing, 1.5 million sharpened factoids have been obtained[2] for 435 sampled sharpened factoids, about 60% were judged reasonable if based on reasonable unsharpened factoids (otherwise the figure was about 40%).

Gleaning conditional rules from discourse cues This very recent work, with preliminary results reported in (Gordon and Schubert 2011), uses sentential patterns containing certain cue words to directly construct conditional (if-then) rules. The technique begins with the application of TGrep to parsed sentences to find instances of patterns such as the following (simplified for readability):

NP VP but didn't VP ,
NP VP, expecting to VP
NP BE ADJP {but|yet} ADJP,

i.e., where an expectation is implied (and perhaps denied). Rules are then applied to the matched sentences, creating slightly simplified and abstracted conditional statements, expressed as parse trees (not yet as LFs).

For example, the sentence

"He stood before her in the doorway, evidently expecting to be invited in"

leads to the rule

If a male stands before a female in the doorway
then he may expect to be invited in.

A few more sample rules are:

If a person texts a male, then he-or-she may get a reply;
If a pain is great, then it may not be manageable;
If a person doesnt like some particular store,
then he-or-she may not keep going to it.

What is interesting about these rules is that they typically posit a consequence relation between distinct event types, described sententially. This is quite different from knowledge acquisition via distributional similarity or via extraction patterns, which typically just relate pairs of lexical items.

About 1 out of 200 sentences yields a rule (that survives filtering); e.g., 29,000 rules were obtained from a 5.5 million sentence personal story corpus. Of these more than 2/3 are judged to be reasonable (based on initial evaluation of a random sample independent of the development set). Conversion to well-formed EL formulas remains to be undertaken, but should be relatively straightforward.

This concludes our survey of the knowledge accumulation methods we have explored so far. As was seen, the abundant factoids we obtained using Knext are for the most part not directly usable for inference, though they are potentially applicable to guiding parser choices. On the other hand, we have also engineered hundreds of lexical axioms, have machine-abstracted 1.5 million quantified, sharpened formulas from the Knext factoids, and most recently obtained many thousands of conditional statements from text. Thus we have begun to test the feasibility of using EPILOG for inference based on the kinds of knowledge we have been accumulating.

5 Implementing NLog (and more) in EPILOG

In this section we move in stages from a consideration of NLog-like inferences in EPILOG to more demanding ones, demonstrating the generality of our approach to knowledge representation and reasoning.

NLog-like inference in EPILOG We noted earlier that NLog entailments are determined by the polarity of environments in which replacements by more general or more specific terms are made. This style of inference fits very well with EPILOG inference, which is also in essence polarity-based: It consists primarily of replacing subformulas by consequences / anti-consequences in positive / negative environments (supplemented by use of natural deduction rules, and, in EPILOG 1, by deployment of specialists).

The second characteristic of NLog, as we mentioned, lies in the way it identifies the complements of implicative and factive verbs (etc.) as entailments or anti-entailments of the sentences as a whole, depending on the way the implicative or factive verbs are embedded. We demonstrated in some of our recent work (Schubert, Van Durme, and Bazrafshan 2010) that we could handle the combination of these strategies quite directly in EPILOG 2. In particular, we showed in some detail how the following illustrative inference used by MacCartney and Manning (2008) could be obtained:

Jimmy Dean refused to move without his jeans
—> James Dean didnt dance without pants.

MacCartney and Manning's NatLog system obtains this inference by first aligning the parsed premise and conclusion with one another as well as possible (bringing identical or similar words into correspondence), and then attempting to transform the premise into the conclusion through a series of local edits (replacement of words or short expressions by more specific or more general ones, depending on polarity, or by entailed or anti-entailed expressions dependent on implicative or factive constructs. Our EPILOG-based demonstration obtained the desired conclusion deductively without having to be told which premise to use. (However, it did have to be told the logical form, rather than the raw English form, of the above premise.)

In our most recent experiments with NLog-like inference, we have focused on deriving entailments based on implicatives, factives, and belief- or want-entailing verbs for a significant number of "natually occurring" examples (Stratos, Schubert, and Gordon 2011). In particular, these experiments were intended to exercise the 250 axioms and inference rules discussed earlier, fueling the EPILOG inference engine. The examples tested included several headlines, such as the following (shown with the corresponding inferences),

- Meza Lopez confessed to dissolving 300 bodies in acid (Examiner: Feb 22, 2011), [Therefore, Meza Lopez dissolved 300 bodies in acid.]
- Oprah is shocked that President Obama gets no respect (Fox News: Feb 15, 2011). [Therefore, Obama gets no respect.]

and 108 sentences randomly selected from the Brown corpus (but restricted to ones containing our target vocabulary of implicative, factive, and belief-/want-entailing lexical items), such as

- I know that you wrote this in hurry. [Therefore, you wrote this in a hurry.]
- They say that our steeple is 162f high. [Therefore, they probably believe that our steeple is 162f high.]

[2] accessible at http://www.cs.rochester.edu/research/knext/browse/

Since there is no full, reliable English to EL parser/interpreter as yet, inputs encoding these sentences for EPILOG were obtained by hand, but guided by outputs of a (rather error-prone) parser/interpreter, based on Dekang Lin's Minipar. Some simplifications were made such as merging some phrases into words by hyphenation, and omitting tense. For example, the logical forms used for the above newspaper headlines were

```
(Meza-Lopez confess
   (ka (l x (some y (y ((num 300) (plur body)))
                    (x dissolve y)))))),
(Oprah (pasv shock) (that (not (Obama get (K respect))))).
```

The conclusions were generated in split seconds (with mapping of both the formalized premise and the conclusion back to English); in an assessment of the validity of the conclusions for the 108 Brown sentences by five judges, 92% were rated as good (75%) or fairly good (17%). The main problem that occurred was just that some sentences were difficult to make sense of out of context, especially when rendered imperfectly back into English from the logical form.

Pushing the limits of NLog: Eventive inferences While the above inferences for headlines and the Brown corpus are all of the type that NLog is designed to handle with ease, inferences based on our growing repertoire of axioms for primitive verbal predicates describing actions and events push or exceed the limits of of current systems for NLog.

For example, we showed an axiom for 'ask-of' in section 3, stating that if x asks y to do something, this conveys to y that x wants y to do it. The entailment relation between asking and conveying a want is not one readily achieved by replacing one subexpression by another, and thus unlikely to be within the capabilities of NLog reasoners such as MacCartney and Manning's NatLog. But in EPILOG 2, when we add to the 'ask-of' axiom a base premise such as "John asked Mary to sing" (neglecting tense),

```
[[John ask-of Mary (Ka sing)] ** E1],
```

the question whether the following is true,

```
[[John convey-info-to Mary
    that [[John want-tbt
            (that (some e2 [e2 right-after E1]
                           [[Mary sing] ** e2]))] @ E1])] * E1],
```

is answered affirmatively by EPILOG in 1 millisecond (in the absence of irrelevant premises in the KB). An important aspect of this inference is its specification of temporal relations among situation types, which is also beyond the scope on current NLog systems.

To date, we have conducted only a small-scale test – 10 questions requiring the above sort of eventive inferences for 5 primitives (ask-of, attack, become, convey-info-to, and know-tbt). But in all cases, entailed conclusions are verified very quickly. For example, given that John became hungry, the inference is made that John was not hungry at the beginning of the becoming-hungry episode (4ms), and was hungry at the end of it (6ms). The most complex input tested was that "Alice conveyed to Bob that Cheryl was asleep", allowing confirmation that Bob knew at the end of the episode that Cheryl was asleep (30ms), that Bob believed at the end of the episode that Cheryl was asleep (34ms), and that Cheryl was in fact asleep during the information-conveying action (24ms).

Beyond the limits: using world knowledge One of the examples in (Stratos, Schubert, and Gordon 2011) combined NLog-like and world knowledge-based inference, but the example was constructed for illustrative purposes. We have recently begun to take some more realistic examples from the "Monroe domain" (emergency response) dialogues collected by James Allen and his collaborators (Stent 2000). While space limitations do not allow detailed presentation of examples, the outline of one of the inferences is as follows:

Every available crane can be used to hoist rubble onto a truck.
Every device that is not in use is available.
Every crane is a device.
The small crane, which is on Clinton Ave, is not in use.

Therefore, the small crane can be used to hoist rubble from the collapsed building on Penfield Rd onto a truck.

Given the premises (in EL form, but neglecting episodic variables), EPILOG 2 confirmed the conclusion in .127 seconds. While the reasoning involves some simple entailment inferences of the type handled in NLog, it involves much else. In fact, in view of the intensional aspects of the first premise, this problem would severly challenge any inference system we are aware of.

A problem that required more extensive world knowledge, and as a result occupied EPILOG for 4 seconds, was roughly as follows:

Most of the heavy resources are in Monroe-east.
[World knowledge about mutual exclusion and joint exhaustion of Monroe-east and Monroe-west, knowledge that 'heavy' is subsective, that if most P are not Q then few P are Q, etc.]

Therefore, there are few heavy resources in Monroe-west.

Note the direct handling of vague nonstandard quantifiers, which again would be a severe challenge to most extant inference systems.

6 Concluding remarks

We have provided additional evidence (beyond that in prior publications) in the form of a broad range of examples that the EL / EPILOG framework directly allows for NL-like representation of world knowledge and lexical knowledge, including lexical schemas (meta-axioms) and rules for Nlog-like inference, but not limited to these.

We have also provided evidence for substantial progress in the acquisition of both lexical and world knowledge. The current number of lexical axioms is still in the hundreds, but our systematic, WordNet- and VerbNet-exploiting methods of adding to these axioms should allow rapid advance on this front. Also we have prepared the ground in past work for use of resources such as Cyc. On the world knowledge side, we have acquired hundreds of millions of shallow general factoids, of predominantly reasonable quality. In efforts to refine this knowledge, we have demonstrated the feasibility of deriving quantified selectional preference-like formulas from clusters of related factoids, as well as the feasibility of sharpening individual factoids with with semantically informed pattern-transduction rules, leading concretely to 1.5 million quantified, inference-capable conditional formulas.

Further, new methods of mapping sentences containing discourse cues have yielded preliminary versions of thousands of if-then rules typically relating two possibly fairly complex clauses concerning distinct kinds of events or properties.

We have also shown through our evolving clusters of example inferences, some evaluated using human judges, that EPILOG 2 makes effective use of sometimes quite complex, NL-like premises, including not only those needed to demonstrate NLog-like entailment inferences, but also ones that encode miscellaneous types of knowledge about the world, often involving relations among events or situations.

There are not many projects similar to our own. James Allen and Johan Bos are among those pursuing ambitious programs aimed at broad-coverage language understanding (e.g., (**?**; Bos 2000)). The currently dormant Bridge system (Bobrow et al. 2007) at Xerox PARC as well was aimed at broad text processing and question answering. The main difference from our approach is the choice of representational and inference frameworks. Allen maps into a thematic role-based representation that allows application of tools such as OWL and his interval temporal logic. Bos maps to DRT and thence to FOL, citing the advantages of being able to employ effective FOL theorem provers. Bridge, like Allen's approach, maps to a thematic role-based representation, and uses light inference methods based primarily on simple entailments and contradiction. Our approach is distinctive in its adherence to an NL-like representation (EL) and use of inference methods subsuming FOL and directly usable for that representation. The recent surge of interest in NLog, which can be seen as a move towards a representation that is essentially parsed NL, and which is energized by the ease of performing many obvious inferences in such a representation, provides new evidence that NL-like representations deserve further exploration and development in the pursuit of human-level NLP and inference.

In knowledge acquisition, our work also runs parallel with some similar efforts, such as were reported in (Sekine 2008). For the most part, however, these related efforts are not aimed at obtaining formalized, inference-capable knowledge, though for instance the ISP system (Pantel et al. 2007) is intended to provide if-then connections between binary relation, with argument type constraints (often the connections are subsumption or similarity of some sort), and Schoenmackers et al. (2010) derive Horn clauses from tuples found by Textrunner (Banko et al. 2007). The general rules learnable in this way seem to be quite different from those our methods produce, and in that sense complementary. Conversely the rules we obtain by factoid sharpening, discourse cue exploitation, and of course knowledge engineering seem beyond the scope of methods employed in these parallel efforts.

One may well ask why we do not directly interpret general statements such as may be found in WordNet glosses, the MIT Open Mind Common Sense project (Havasi, Speer, and Alonso 2007), or Wikipedia. We consider this a proximate goal, but it faces the many difficulties that still beset the mapping from NL to logical form, even one as close to NL as EL. For example, the WordNet gloss for the verb 'dance' is

dance (V): move in a pattern, usually to musical accompaniment

But what does "in a pattern" mean? (Cf. "move into / inside a pattern".) And what does "to musical accompaniment" mean? (towards the accompaniment?) As another sort of example, a fact found in Open Mind about car crashes is

Something you might do while driving a car is crash,

but this leaves entirely open who/what is crashing into what, let alone the temporal relation between the driving and the crashing. It is sometimes suggested that NLog finesses many of the ambiguities of language by operating directly on language. While we have indicated that NLog is indeed somewhat tolerant of vagueness and ambiguity, real language understanding cannot leave in place indexicals (I, you, now, this, ...) and anaphors, whose semantic value is context- and time-dependent, nor can it abide radical word sense ambiguity. For example, "John had gerbils as a child" should not be taken to entail that John ate, or gave birth to, small rodents as a child – as might well be decided by an NLog system. So predicate disambiguation, as well as modifier attachment, de-indexing, operator scoping, coreference, implicit argument detection, temporal analysis, and other issues still await solution in an integrated way before we can expect machines to truly understand us, make inferences like us, and learn by reading.

Acknowledgements

The authors gratefully acknowledge Fabrizio Morbini's vital and expeditious help in extending EPILOG 2 for our purposes and resolving various issues. Also, thanks to Ben Van Durme for his important earlier contributions to the KA effort, which are still reflected here. The work was supported by NSF grants IIS-1016735 and IIS-0916599, and a subcontract to ONR STTR contract N00014-10-M-0297.

References

Banko, M.; Cafarella, M. J.; Soderland, S.; Broadhead, M.; and Etzioni, O. 2007. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, 2670–2676.

Bobrow, D. G.; Cheslow, B.; Condoravdi, C.; Karttunen, L.; Holloway King, T.; Nairn, R.; de Paiva, M.; Price, C.; and Zaenen, A. 2007. PARCs Bridge and Question Answering System. In Holloway King, T., and Bender, E. M., eds., *Proceedings of the GEAF 2007 Workshop, CSLI Studies in Computational Linguistics ONLINE*. CSLI Publications.

Bos, J. 2000. Wide-coverage semantic analysis with Boxer. In *Symposium on Semantics in Systems for Text Processing (STEP 2008) Shared Task: Comparing Semantic Representations*, 277286.

Danescu-Niculescu-Mizil, C.; Lee, L.; and Ducott, R. 2009. Without a 'doubt'? unsupervised discovery of downward-entailing operators. In *Proc. of NAACL HLT*, 137–145.

Gordon, J., and Schubert, L. K. 2010. Quantificational sharpening of commonsense knowledge. In *Proc. of the AAAI 2010 Fall Symposium on Commonsense Knowledge (CSK10)*.

Gordon, J., and Schubert, L. K. 2011. Discovering commonsense entailment rules implicit in sentences. In *Proc. of the EMNLP 2011 Workshop on Textual Entailment (TextInfer 2011)*.

Gordon, J.; Van Durme, B.; and Schubert, L. K. 2010. Learning from the web: Extracting general world knowl-

edge from noisy text. In *Proc. of the AAAI 2010 Workshop on Collaboratively-built Knowledge Sources and Artificial Intelligence (WikiAI)*.

Havasi, C.; Speer, R.; and Alonso, J. 2007. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. In *Recent Advances in Natural Language Processing*.

Kaplan, A. N., and Schubert, L. K. 2001. Measuring and improving the quality of world knowledge extracted from WordNet, tech. rep. 751. Technical report, Dept. of Computer Science, Univ. of Rochester, Rochester, NY, USA.

MacCartney, B., and Manning, C. D. 2008. Modeling semantic containment and exclusion in natural language inference. In *Proc. of the 22nd International Conference on Computational Linguistics (COLING '08)*.

Morbini, F., and Schubert, L. K. 2008. Metareasoning as an integral part of commonsense and autocognitive reasoning. In Cox, M. T., and Raja, A., eds., *Metareasoning: Thinking about Thinking*.

Nairn, R.; Condoravdi, C.; and Karttunen, L. 2006. Computing relative polarity for textual inference. In *Inference in Computational Semantics (ICoS-5)*, 20–21.

Pantel, P.; Bhagat, R.; Chklovski, T.; and Hovy, E. 2007. Isp: Learning inferential selectional preferences. In *In Proceedings of NAACL 2007*.

Schoenmackers, S.; Davis, J.; Etzioni, O.; and Weld, D. S. 2010. Learning first-order horn clauses from web text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*, 1088–1098.

Schubert, L. K., and Hwang, C. H. 2000. Episodic Logic Meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In Iwanska, L., and Shapiro, S. C., eds., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*. MIT/AAAI Press.

Schubert, L. K., and Tong, M. H. 2003. Extracting and evaluating general world knowledge from the Brown corpus. In *Proc. of the HLT-NAACL Workshop on Text Meaning*.

Schubert, L. K.; Van Durme, B.; and Bazrafshan, M. 2010. Entailment inference in a natural logic–like general reasoner. In *Proc. of the AAAI 2010 Fall Symposium on Commonsense Knowledge (CSK10)*.

Schubert, L. K. 2002. Can we derive general world knowledge from texts? In *Proc. of the 2nd International Conference on Human Language Technology Research (HLT02)*.

Sekine, S., ed. 2008.

Stent, A. 2000. The monroe corpus, tech. rep. no. tr728 and tn99-2. Technical report, Dept. of Computer Science, Univ. of Rochester, Rochester, NY, USA.

Stratos, K.; Schubert, L.; and Gordon, J. 2011. Episodic Logic: Natural logic + reasoning. In *International Conference on Knowledge Engineering and Ontology Development*.

Van Durme, B.; Michalak, P.; and Schubert, L. K. 2009. Deriving generalized knowledge from corpora using WordNet abstraction. In *Proc. of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*.

Van Durme, B.; Qian, T.; and Schubert, L. K. 2008. Class-driven attribute extraction. In *Proc. of the 22nd International Conference on Computational Linguistics (COLING-08)*, 921–8.