NLog-like Inference and Commonsense Reasoning

Lenhart Schubert[1]

Recent implementations of Natural Logic (NLog) have shown that NLog provides a quite direct means of going from sentences in ordinary language to many of the obvious entailments of those sentences. We show here that Episodic Logic (EL) and its Epilog implementation are well-adapted to capturing NLog-like inferences, but beyond that, also support inferences that require a combination of lexical knowledge and world knowledge. However, broad language understanding and commonsense reasoning are still thwarted by the "knowledge acquisition bottleneck", and we summarize some of our ongoing and contemplated attacks on that persistent difficulty.

1 Introduction: Natural Logic, Episodic Logic, and Epilog

Natural Logic (NLog, as we will abbreviate it) provides an attractive framework for implementing many "obvious" entailment inferences, starting with little more than a syntactically analyzed sentence (e.g., MacCartney and Manning 2008). But in itself, NLog is certainly insufficient for language understanding or commonsense inference. We advocate here a formalized (not necessarily fully disambiguated) representation that is very close to language, along with an inference methodology that easily accommodates NLog inferences, but beyond that enables more general reasoning based on both lexical knowledge and world knowledge. This approach to representation and inference has been pursued by our group at the University of Rochester (and

---

[1]University of Rochester

prior to that, at the University of Alberta) for over two decades, and the outcome so far is the Episodic Logic (EL) representation and its implementation in Epilog systems 1 and 2. We have also built up very substantial amounts of general knowledge,[2] much of it rough, but a good deal of it precise enough for Epilog-based inference.

In the next section we review the EL representation and the Epilog inference architecture. In section 3 we illustrate NLog-like inferences in Epilog, and touch on some evaluation results for such inferences. Section 4 provides examples of commonsense inferences performed by Epilog that lie beyond the capabilities of NLog, in part because they involve radical differences in phrase structure between the premises and the conclusion and in part because they draw on world knowledge as well as lexical semantic knowledge. However, enhanced inference capabilities alone cannot take us very far towards achieving broad understanding and commonsense reasoning: We still face the knowledge acquisition bottleneck, and in section 5 we discuss multiple facets of our efforts to provide broad knowledge to Epilog. In section 6 we assess the prospects for general knowledge acquisition through genuine understanding of WordNet glosses, Simple Wikipedia entries, or the Open Mind collection of general statements. We sum up the relationship between NLog and Epilog in section 7, also briefly reiterating the status of our knowledge acquisition efforts and the prospects for moving beyond them through deeper understanding of general statements.
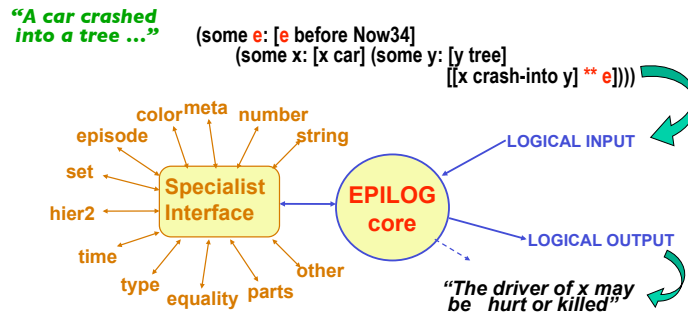
2   EL and Epilog

Figure 1 provides a glimpse of the EL representation and the Epilog architecture. Note the infixing of predicates like *car* and *crash-into* in formulas, used for readability.[3]

Note also the association of an episodic variable e with the subformula [x *crash-into* y], via the operator '**'. Intuitively, the operator expresses that the sentence *characterizes* episode e as a whole (rather than merely some temporal portion or aspect of it). (Schubert and Hwang 2000) provides a general explanation of EL, and (Schubert 2000) explicates the logic of '**' viewed as an extension of first-order logic (FOL), relating this aspect of EL to various extant theories of events and situations (such as those of Davidson, Reichenbach, and Barwise & Perry). Given an input such as is shown in the figure, and a general

---

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

[3]As a further aid to readability, printed EL infix formulas are often written with square brackets (while prefix expressions use round brackets), and restricted quantification is indicated with a colon after the variable. In the "computerized" version, all brackets are round and there are no colons preceding restrictors.

FIGURE 1: Episodic Logic and the EPILOG system

axiom that if a car crashes into a tree, its driver may be hurt or killed, EPILOG would draw the expected conclusion. This is of course an inference based on world knowledge, not one feasible in NLog systems, at least as so far developed.

EPILOG 1 performs forward (input-driven) as well as backward (goal-driven) inference, and is fully integrated with the indicated specialist modules. However, its selective knowledge retrieval method and somewhat patchwork construction (driven by a succession of contracts) leave it with certain hard-to-predict blind spots in its backward inference, and these are remedied in the more recent EPILOG 2 version.[4]

The EL language is Montague-inspired and is structurally and semantically very close to natural language (NL). We might say that it is scope-resolved, partially disambiguated NL with variables, including explicit episodic (i.e., event and situation) variables. Historically, most research on NL understanding and dialogue has taken a depth-first approach, creating end-to-end systems for highly focused problem solving or text understanding domains, and trying to build "outward" from these. We have instead pursued a breadth-first approach, attempting

---

[4]Implemented by University of Rochester PhD graduate Fabrizio Morbini, now at the Institute for Creative Technologies in California. Forward inference is only partially implemented in EPILOG 2 at this point, and the specialists are not yet re-integrated.

to build general frameworks for representing knowledge (especially verbalizable knowledge), semantic interpretation, and inference; we wished to avoid the temptations and pressures that arise in specialized applications to use domain-specific representations and interpretive heuristics that are unusable outside the chosen domain.

There are various considerations that make the idea of a language-like internal ("Mentalese") meaning representations plausible:

- A common conjecture in anthropology and cognitive science is that language and thought appeared more or less concurrently, perhaps some 200,000 years ago.[5] This makes it likely that they are variants of the same basic symbolism; after all, language serves to communicate thoughts.

- The assumption that what comes out of our mouths closely resembles what is in our heads (at an appropriately abstract level of analysis, and allowing for time-saving abbreviations, omissions, and other "telegraphic" devices) is prima facie simpler than the assumption that the two representational systems diverge radically; especially so if we keep in mind that NL understanding is incremental, requiring that fragmentary interpretations be brought into inferential contact with stored knowledge.

- All our symbolic representations, at least those based on tree-structured expressions, are derivative from language; this includes logics, programming languages, semantic nets, and other AI-oriented knowledge representations. This suggests a language-like substrate in our higher-level thinking.

- Surely, it cannot be a mere coincidence that, as shown by Montague, entailment can be understood in terms of semantic entities corresponding one-to-one with syntactic phrases.

- Recent successes in applying NLog to entailment inference underscore the advantages of working directly with (structurally analyzed) NL.

These considerations may strike many linguistic semanticists (and perhaps cognitive scientists and anthropologists) as belaboring the familiar idea of "language as a mirror of mind". But in AI there is strong resistance among many researchers to extending the expressivity of knowledge representations to full first-order logic, let alone *beyond* FOL

---

[5]Preceded, perhaps, by around 2 million years of evolution of the mirror neuron system and Broca's area, allowing imitation of the actions and gestures of others, and grasp of their intentions (Fadiga et al. 2006). Recent fMRI studies suggest that a distinct portion of Broca's area is devoted to cognitively hard tasks such as arithmetic (Fedorenko et al. 2012).

– despite the difficulty of mapping many easily verbalized ideas into FOL, frames, or description logics. The argument made is that expressivity must be reined in to achieve inferential efficiency and completeness. Though we won't elaborate here, this seems to us a false trade-off, much like denying programmers the ability to use recursion or looping (with the number of iterations as a run-time variable), to prevent them from writing inefficient programs. Greater expressivity allows not only greater coverage of ideas expressible in language but also more efficient inference in cases where weakly expressive representations would require complex work-arounds, if feasible at all.

The goal in the design of EL was to match the expressive devices shared by all human languages. These include the following:

- Ways of naming things
- Boolean connectives: and, or, not, if-then, ...
- Basic quantifiers: every, some, no, ...
- Ways of ascribing properties and relationships to entities
- Identity

These items alone already imply the expressivity of FOL, at least if we do not impose arbitrary restrictions, for instance, on quantifier embedding or predicate adicity. But natural languages allow more than that:

- Generalized quantifiers (Most women who smoke)
- Intensionality (is planning a heist; resembles a Wookiee)
- Event reference (Members of the crowd hooted and shouted insults; *this* went on for minutes)
- Modification of predicates and sentences (*barely* alive, dances *gracefully*, *Perhaps* it will rain)
- Reification of predicates and sentences (*Xeroxing money* is illegal; *That there is water on the Moon* is surprising)
- Uncertainty (It will *probably* rain tomorrow; The more you smoke, the greater *your risk of developing lung cancer*)
- Quotation and meta-knowledge (Say "cheese"; How much do you know about description logics?)

All of these devices are directly enabled in EL. Here are some illustrative examples of logical forms:

- Restricted quantifiers
  "*Most laptops are PCs or MACs*"
  ```
  [<Most laptop> <or PC MAC>]            {unscoped LF}
  (Most x: [x laptop] [[x PC] or [x MAC]]) {scoped LF}
  ```

- Event relations
  "*If a car hits a tree, the driver is often hurt or killed*"

```
(Many-cases e:
  (some x: [x car] (some y: [y tree] [[x hits y] ** e]))
  [[[(driver-of x) (pasv hurt)]  or
     [(driver-of x) (pasv kill)]] @ e])
```

  The episodic operator '@' is a variant of '**', expressing that a sentence characterizes some episode *concurrent with* a given episode *e*.

- 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])
```

Note the predicate modifiers *firmly, plur*, and *nearly*, and the reifying operators *that* and *K*; *that* maps sentence intensions (partial functions from possible episodes to truth values) to individuals, and *K* maps predicate intensions to individuals, namely *kinds* in the sense of (Carlson 1977). Further features are the allowance for quoted syntactic expressions as terms, and substitutional quantification over expressions of all sorts. In this way Epilog can entertain propositions about syntactic entities such as names and other linguistic expressions, telephone numbers, mathematical expressions, Lisp programs, or its own internal formulas, and can use formalized axiom schemas for inference – a crucial capability for reasoning with general meaning postulates in NLog-like manner, as we will see. The following example illustrates the use of substitutional quantification and (quasi)quotation to express two claims about name-knowledge:

"*I know the names of all CSC faculty members*"

```
(all x: [x member-of CSC-faculty]
       (all_subst y: ['y name-of x]
                       [ME know (that ['y name-of x])])))
```

"*There is no CSC faculty member whose name I know to be
'Alan Turing'.*"

```
(no x: [x member-of CSC-faculty]
       [ME know (that ['(Alan Turing) name-of x])])
```

From this Epilog would easily infer that there is no member of the CSC faculty named Alan Turing. Incidentally, contrary to the notion that high expressivity entails low efficiency, experimental application of Epilog to large-scale theorem proving in first-order logic showed it to be competitive with the best theorem provers, especially for relatively

large axiom bases (Morbini and Schubert 2009).

3   NLog-like inference in Epilog

The essential ideas behind NLog are the following (e.g., van Benthem 1991, 2007; Valencia 1991; van Eijck 2005; Nairn et al. 2006; MacCartney and Manning 2008):

1. Starting with a syntactically structured natural language sentence, we can replace phrases by more general [more specific] ones in positive- [negative-] polarity environments; identity or equivalence replacements are permissible in both types of environments as well as in (transparent) environments that are neither upward nor downward entailing;

    e.g., *Several trucks are on their way* →
    Several vehicles are on their way;
    If a vehicle is on its way, turn it back →
    If a truck is on its way, turn it back

2. We exploit implicatives/factives;

    e.g., *X manages to do Y* → *X do Y*;
    X doesn't manage to do Y ⇝ *X doesn't do Y*;
    X knows that Y → *Y*
    X doesn't know that Y → *Y*;

3. Full disambiguation is not required; e.g., *several* and *on their way* in (1) above can remain vague and ambiguous without disabling the indicated inferences.

Like NLog inference, inference in Epilog is *polarity-based*. In essence, it consists of replacing subformulas in arbitrarily complex formulas by consequences or anticonsequences in positive and negative polarity environments, respectively (and using identity or equivalence substitutions much as in NLog).

The equivalents of Nlog inferences are readily encoded via axioms and rules in Epilog 2. For example, we have duplicated MacCartney and Manning's illustrative example,

Jimmy Dean refused to move without his jeans
    → *James Dean didn't dance without pants*

However, in Epilog the (anti)consequences may depend on world knowledge as well as lexical knowledge; also polarity-based inference is supplemented with natural deduction rules, such as assumption of the antecedent in proving a conditional, or reasoning by cases in proving a disjunction.

A restriction in Epilog is that it replaces only sentential parts of larger formulas, while NLog can replace fragments of arbitrary types.

However, this makes little practical difference. For example, NLog might replace *truck* by *vehicle* in a positive environment, while Epilog would replace a formula of form [*τ truck*] by [*τ vehicle*]. The result is the same, at least when *τ* is unaltered. (An example where both the predicate and its argument are altered would be in the replacement of [*Sky overcast*] by [*Weather cloudy*].) Another example would be the replacement of *many* by *some* in NLog, while Epilog would replace a sentence of form (*many α: φ ψ*) by one of form (*some α: φ ψ*). Again, the effect is the same in cases where the operands of the quantifier are unaltered. Note that the *truck/vehicle* example above would depend in Epilog on an ordinary axiom, (*all x*: [*x truck*] [*x vehicle*]), while the *many/some* example would depend on an axiom *schema*, i.e., one that quantifies susbstitutionally over formulas (in the restrictor and nuclear scope).

The replacements effectuated by Epilog in performing NLog-like inferences based on implicatives also typically depend on axiom schemas. The following pair of schemas for the implicative *dare* illustrate this point:

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

These schemas capture the fact that if someone dared to do something, they did it, and if they didn't dare do something, they didn't do it. We can say that the *signature* of *dare* is of type $+/-$, to indicate its positive entailment in a positive environment and its negative entailment in a negative environment. Here *Ka*, the logical counterpart of the infinitive particle *to*, is another predicate reifying operator, forming a *kind of action or attribute* when applied to verbal predicates. (Actions are treated in EL as consisting of an individual – the agent or subject of the predication – and an episode. *Ka* is definable in terms of *K* and the episodic operator '**'.)

Similar axiom schemas can be provided for other implicatives, such as the following (in stylized rather than precise form):

> *X decline to P => X not P*
> *X not decline to P => (probably) X P*[6]
> *X agrees to P => (probably) X P*
> *X does not agree to P => (probably) not X P*
> *X doubts that W => X believes probably not W.*

The signatures here are $-/(+)$ and $(+)/(-)$ for *decline* and *agree* respectively. Note that some entailments are weakened to implicatures, e.g., it is possible that Bob did not decline to review a certain paper,

---

[6]Whether the qualifier *probably* should be present or not is debatable

yet failed to review it. The example of *doubt* is included here to indicate that certain attitudinal entailments not normally regarded as falling under the implicative or NLog motif can be captured via axiom schemas like those supporting NLog inference.

Besides various axiom schemas, matching the capabilities of NLog also requires special inference rules. (These are formalized in Epilog much like axiom schemas.) In particular, *factive* (*presuppositional*) verbs such as *know* and *realize*, with $+/+$ signatures in simple assertional contexts, can be partially handled with rules such as

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

where the long arrow indicates inferrability. Clearly we could not use axiomatic versions of these rules, since by the law of excluded middle, either x knows that w or x does not know that w, regardless of x and w, and so axioms stating that w holds in either case would lead us to conclude that w holds, regardless of its content. However, these rules only address part of the "projection problem" for presuppositions. For example, they cannot be applied to the consequent clause of a sentence such as "*If Bob was telling the truth, then Alice knows that he was telling the truth*", which does not entail that Bob was telling the truth. A full treatment of presuppositional clauses, whether in NLog, Epilog or any other framework, would need to take account of the context established by prior clauses in the discourse. In the preceding example, the *if*-clause establishes a context where it is an open question whether Bob was telling the truth, blocking presupposition projection from the *then*-clause.[7]

The following examples are some (rather eye-catching) headlines collected by Karl Stratos, and rendered into EL by him to demonstrate inferences based on implicatives. A significant point to note is that human readers seem to (unconsciously) generate and internalize the corresponding inferences; this point stands in contrast with the current emphasis in NLP on *recognizing* textual entailment, i.e., on making boolean judgements where both the premise and the hypothetical conclusion are at hand, so that piecemeal alignments and transformations as well as statistical guesswork can be employed in the task.

---

[7]As we noted in (Stratos et al. 2011), Clausen and Manning (2009) proposed a way of projecting presuppositions in NLog in accord with the plug-hole-filter scheme of Karttunen (1973). In this scheme, plugs (e.g., 'say') block all projections, filters (e.g., 'if-then') allow only certain ones, and holes (e.g., 'probably') allow all. But the approach does not fully handle the effects of discourse context.

- *Vatican refused to engage with child sex abuse inquiry* (The Guardian: Dec 11, 2010).
- *A homeless Irish man was forced to eat part of his ear* (The Huffington Post: Feb 18, 2011).
- *Oprah is shocked that President Obama gets no respect* (Fox News: Feb 15, 2011).
- *Meza Lopez confessed to dissolving 300 bodies in acid* (Examiner: Feb 22, 2011)

Corresponding EL representations for EPILOG inference are the following (with various contractions for readability, and neglecting tense and thus episodes; (l y ...) indicates lambda-abstraction):

```
[Vatican refuse (Ka (engage-with Child-sex-abuse-inquiry))]
(some x: [x (attr homeless (attr Irish man))]
        [x (pasv force)
           (Ka (l y (some r: [r ear-of y]
                            (some s: [s part-of r] [y eat s]))))])
[Oprah (pasv shock) (that (not [Obama get (K respect)]))]
[Meza-Lopez confess
   (Ka (l x (some y: [y ((num 300) (plur body))] [x dissolve y])))].
```

Obvious inferences confirmed by EPILOG using implicative axiom schemas (in fractions of a second) were the following (returned in English):

The Vatican did not engage with child sex abuse inquiry.
An Irish man did eat part of his ear,
President Obama gets no respect, and
Meza Lopez dissolved 300 bodies in acid.

In addition to treating these illustrative examples, Stratos also extracted a random test set of 108 sentences from the Brown corpus, restricting these to ones containing any of the 250 verbs (including some short phrases) covered by an axiomatic knowledge base. The knowledge base comprised a superset of the collections of implicative and factive verbs obtained from (Nairn et al. 2006), (Danescu-Niculescu-Mizil et al. 2009), and Cleo Condoravdi (personal communication), and also included axioms for separately collected attitudinal verbs, for testing inference of beliefs and desires. Two sample sentences, along with the conclusions drawn from their logical forms, are the following:

I know that you wrote this in a hurry.
    → *You wrote this in a hurry.*

They say that our steeple is 162ft high.
    → *Probably they believe that our steeple is 162ft high.*

Note that the latter inference falls outside the scope of standard NLog inference, since it involves not only a substitution (of *believe* for *say*) but

also a simultaneous premodifier insertion (*probably*). The logical forms of the 108 sentences (obtained by manual correction of flawed outputs from the current NL-to-EL interpreter) led to 141 inferences rendered automatically into English, which were rated by multiple judges. They were predominantly judged to be good (75%) or fairly good (17%). Lower ratings were due mostly to incomprehensibility or vacuity of the conclusion, as in "*The little problems help me to do so*" → "*I do so*". Further details can be found in (Stratos et al. 2011).

Our conclusion is that NLog inferences are readily implemented within the EL-Epilog framework, though further work on the NL-to-EL interface is needed for full automation. (The errors come mostly from faulty parses, which are certainly an issue in automating NLog as well.)

4 Commonsense inferences beyond the scope of NLog

Not all lexical entailments and implicatures are as simple as those we have focused on so far. For example, asking someone to do something entails conveying to them that one wants them to do it. This is expressed in EL with the schema

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

(Here '*' is weakening of '**' from characterization to partial description – see Schubert 2000.) Thus, given that "*John asked Mary to sing*",

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

the question "*Did John convey to Mary that he wanted her to sing?*",

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

is immediately answered in the affirmative by the Epilog system.

Even further removed from the current reach of NLog are inferences dependent on world knowledge along with lexical knowledge. The premises in the following desired inference are based on a (human-to-human) dialogue excerpt from James Allen's and George Ferguson's Monroe domain (Stent 2000), where the dialogue participants are considering what resources are available for removing rubble from a collapsed building. The first premise is presumed background knowledge,

and the second premise reflects an indirect suggestion by one of the participants:

Every available crane can be used to hoist rubble onto a truck.
The small crane, which is on Clinton Ave, is not in use.
→ The small crane can be used to hoist rubble from the collapsed building on Penfield Rd onto a truck.

The two premises are expressed in EL as follows:

```
(all x: [x ((attr available) crane)]
  (all r: [r rubble]
    [(that (some y: [y person]
            (some z: [z truck]
              [y (adv-a (for-purpose (Ka (adv-a (onto z) (hoist r))))
                        (use x))]))) possible]))

(the x: [x ((attr small) crane)]
        [[x on Clinton-Ave] and (not [x in-use])])
```

We also need to make a connection between the notion of a device not being in use and being available, [8] and affirm that cranes are devices:

```
(all x: [[x device] and (not [x in-use])] [x available])
(all x: [x crane] [x device])
```

If we now pose the desired inference as the question, "*Can the small crane be used to hoist rubble from the collapsed building on Penfield Rd onto a truck?*",

```
(the x: [x ((attr small) crane)]
  (some r: [[r rubble] and
            (the s: [[s (attr collapsed building)] and
                     [s on Penfield-Rd]] [r from s])]
    [(that (some y: [y person]
            (some z: [z truck]
              [y (adv-a (for-purpose (Ka (adv-a (onto z) (hoist r))))
                        (use x))]))) possible]))
```

an affirmative answer is produced by Epilog (in about 1/8 sec on a Dell workstation with an Intel dual core CPU @ 2GHz; however, the time would go up with a larger knowledge base). Such inferences are made by humans just as quickly (and unconsciously) as those based on implicatives, but they are not possible in NLog as currently understood.

Two related inference examples from the same domain are the following.

Most of the heavy Monroe resources are located in Monroe-east.
→ Few heavy resources are located in Monroe-west.
→ Not all Monroe resources are located in Monroe-west.

---

[8] In the Monroe domain, and perhaps in task-oriented dialogues more broadly, the phrase *not in use* (applied to equipment) pretty reliably indicates *available for use*.

These depend on knowledge about the meanings of vague quantifiers, and on knowledge about the way in which we often conceptually partition geographical regions into disjoint eastern, western, northern, and southern parts. The following are some natural lexical schemas and geographical axioms:

- *If most P are not Q, then few P are Q*:
  ```
  (all_pred P
    (all_pred Q
      [(most x: [x P] (not [x Q])) => (few x: [x P] [x Q])]))
  ```
- *"Heavy" in premodifying position is subsective*:
  ```
  (all_pred P (all x: [x ((attr heavy) P)] [x P]))
  ```
- *If most P are Q, then some P are Q* (existential import of "most"):
  ```
  (all_pred P
    (all_pred Q
      [(most x: [x P] [x Q]) => (some x: [x P] [x Q])]))
  ```
- *All Monroe resources are in Monroe. A thing is in Monroe iff it is in Monroe-east or Monroe-west; and iff it is in Monroe-north or Monroe-south; nothing is in both Monroe-east and Monroe-west; or in both Monroe-north and Monroe-south*:
  ```
  (all x: [x Monroe-resources] [x loc-in Monroe])
  (all x: [[x loc-in Monroe] <=>
      [[x loc-in Monroe-east] or [x loc-in Monroe-west]]])
  (all x: [[x loc-in Monroe] <=>
      [[x loc-in Monroe-north] or [x loc-in Monroe-south]]])
  (all x: [(not [x loc-in Monroe-east]) or (not [x loc-in Monroe-west])])
  (all x: [(not [x loc-in Monroe-north]) or (not [x loc-in Monroe-south])])
  ```
- *There are some heavy Monroe resources; most of the heavy Monroe resources are located in Monroe-east*:
  ```
  (some x [x ((attr heavy) Monroe-resources)])
  (most x: [x ((attr heavy) Monroe-resources)] [x loc-in Monroe-east])
  ```

Some questions that can now be tackled are the following (with a *yes* and *no* answer respectively):

Are few heavy resources in Monroe-west?

```
(few x: [x ((attr heavy) Monroe-resources)] [x loc-in Monroe-west])
```

Are all Monroe resources in Monroe-west?

```
(all x: [x Monroe-resources] [x loc-in Monroe-west])
```

The reasoning behind the *no* answer to the latter question is this: Most heavy resources, hence some heavy resources, hence some resources, are in Monroe-east; but whatever is in Monroe-east is not in Monroe-west, hence not all resources are in Monroe-west. These reasoning examples took a few seconds, probably because EPILOG 2's limited form of input-driven inference did not fire in this case. But one may also speculate

that a spatial reasoning specialist could accelerate inference here – people seem to build or access a mental model of the spatial layout alluded to in such examples, and "read off" the answers.

5 Tackling the knowledge acquisition bottleneck

As long as our knowledge bases are small, or restricted to approximate relations between pairs of lexical items, we cannot hope to achieve broad language understanding in machines, or come anywhere close to matching the human capacity for spontaneously generating commonsense inferences upon receipt of new inputs. Therefore, we now turn to the long-standing issue of scaling up a general knowledge base to enable understanding and inference for miscellaneous texts.

We divide the knowledge we are seeking into three general sorts: (a) *lexical knowledge* (for NLog-like and other meaning-based inference); (b) *semantic pattern knowledge* (which we have come to regard as a separate category needed to guide parsing and interpretation, and as a starting point for formulating deeper knowledge); and (c) *world knowledge*, essential to our comprehension of and reasoning about everyday entities and events.

While our methods of mining textual resources for general knowledge, and methods of knowledge accumulation pursued by others, have yielded millions of items of formal knowledge, many of them inference-enabling, these methods still seem far too weak to capture the breadth and richness of human commonsense knowledge. We believe that knowledge bootstrapping ultimately must be grounded in actual, deep understanding of a limited, but ever-growing range of sentences. The most direct method of bootstrapping would be to supply large numbers of explicitly stated generalizations to a system, where these generalizations are presented in the form of relatively simple, relatively self-contained statements. Such generalizations should provide much of the "glue" that binds together sentences in coherent discourse, and should allow generation of expectations and explanations. Thus we will follow our discussion of our efforts to acquire knowledge of types (a-c) with a commentary on the remaining hurdles in deriving deep, logically defensible representations of the content of generic sentences in such sources as WordNet glosses, Simple Wikipedia entries, and Open Mind factoids.

5.1 Acquiring lexical semantic knowledge

Our several approaches to acquiring lexical meaning postulates have been motivated by the availablity of several resources, namely distributional similarity clusters, WordNet hierarchies, collections of implicative and factive verbs, and VerbNet classes. In particular we have pur-

sued the following strategies.

Finding relations among similar words: Starting with distributional similarity clusters (made available to us by Patrick Pantel), we used supervised machine learning techniques to identify entailment, synonymy, and exclusion relations among pairs of words in a cluster, using features borrowed from (MacCartney and Manning 2009) such as WordNet distance, the DLin feature based on Dekang Lin's thesaurus, and morphosyntactic features. After transitivity pruning, we formalized these relations as quantified axioms. Accuracies ranging from 65% to over 90% (depending on word class and relation) were attained (Schubert et al. 2010). While this sort of accuracy may be sufficient for statistical entailment judgements, it falls short of the kind of accuracy we are trying to achieve for reasoning purposes. Indeed, the "gold standard" examples used in training were only as reliable as the judgements of the humans responsible for the labeling, and we have come to the conclusion that such judgements need to be refined before they can be usefully deployed for axiom generation. For example, a lexicographer or graduate student in NLP may well judge that *cognition* entails ("isa") *psychological feature*, yet on careful consideration, it makes little sense to say that "*Every entity with the property of being (a?) cognition has the property of being a psychological feature*". Rather, cognition, as a *kind* of activity performed *by some entity* could be said to be a psychological feature *of* that entity. Similarly, while we might casually judge each link in a hypernym chain such as *ellipse → conic section → shape → attribute* to correspond to an entailment, the conclusion that every ellipse is an attribute is unacceptable.

Thus in current work we are focusing on formulating hypotheses concerning the features of synset pairs classified as hyponym-hypernym pairs in WordNet that are indicative of a specific type of logical relation. For example, whereas a hyponym-hypernym transition from (some sense of) a count noun to (some sense of) another count noun is typically (though by no means always) an entailment transition, and similarly for mass-mass transitions, in the case of a mass-count transition we are more likely confronted with an *instance* relation between a *kind* derived from a mass predicate and a count predicate applicable to kinds. For example, in the hypernym chain *gold dust → gold → noble metal → metallic element → element* we can perfectly well infer that every entity with the property of being gold dust has the property of being gold; or, that every entity with the property of being a noble metal has the property of being a metallic element; however, we should not conclude (transitively) that every entity with the property of being gold dust

has the property of being a metallic element. The breakdown occurs at the mass-count boundary between *gold* and *noble metal*: The correct reading of the transition here is not as one from property to property, but as a predication about a kind, viz., the kind, *gold*, is a noble metal. We are currently able to generate axioms from the majority of direct hypernym relations in WordNet, but our casual observation that the great majority of these axioms are tenable awaits formal confirmation.

Unfortunately, hierarchy relations for nouns, though fundamentally important, do not go very far towards capturing their meanings. For example, in the case of gold it is essential to know not only that it is a noble metal, but also that it is mined, that it is of high density and malleable, that it melts at sufficiently high temperatures, that it is prized as material for jewelry, and so on. This observation ties in with our comments below on the need for "object-oriented" knowledge in verb-based inference.

Axiomatizing implicative verbs: As already noted in our reference to Karl Stratos' work, we have undertaken a knowledge engineering effort to collect factive and implicative verbal predicates (along with some antifactive and belief- or want-implying ones) and axiomatize them for use in Epilog. The 250 entries were gleaned from various sources and expanded via VerbNet, thesauri, etc. But as pointed out by Karttunen (2012), there are also numerous *phrasal implicatives*, such as *make a futile attempt to*, *make no effort to*, *take the trouble to*, *use the opportunity to*, or *fulfill one's duty to*. It is doubtful that such cases can be adequately treated by an enumerative approach; rather it appears that multiple items of lexical knowledge will need to be used in concert to derive the desired entailments or implicatures.

Axiomatizing Verbnet classes: After an initial attempt to directly render VerbNet semantic annotations into Epilog axioms, which led to generally weak and often flawed axioms, we switched to a more meticulous approach to acquiring formal verb semantics (as first sketched in (Schubert et al. 2011)). We have been tackling one verb class at a time, typically formulating an axiom schema for the class (or two or three schemas for subsets of the class), and instantiating the parameters of these schemas with particular predicates or modifiers for particular verbs. For example, the VerbNet change-of-state class *other_cos-45.4*, which includes such verbs as *clean, darken, defrost, open, sharpen,* and *wake*, can be partially characterized in terms of an axiom schema that states that in an event e characterized by an animate agent X acting on a physical entity Y in accord with one of these verbs, Y will have a certain property P at the beginning of the event and a (contrasting)

property Q at the end. P and Q might be the (adjectival) predicates *dirty.a* and *clean.a* for verbal predicate *clean*, *light.a* and *dark.a* for *darken*, *frozen.a* and *unfrozen.a* for *defrost*, and so on. (However, quite a few of the 384 verbs, such as *abbreviate, brighten, dilute, improve,* and *soften* require axiomatization in terms of a change on a scale, rather than an absolute transition.) For many classes, the axiom schemas make use of separately axiomatized *primitive* verbal predicates; these number about 150, and were chosen for their prevalence in language, early acquisition by children, inclusion in primitive vocabularies by others, and utility in capturing meanings of other verbs.

This is decidedly still work in progress, with 116 axioms for primitives and 246 for verbs (in 15 VerbNet classes) in the lexical knowledge base so far. A related ongoing project is aimed at evaluation of the axiom base via human judgements of forward inferences based on the axioms. But even completion of these projects will leave untouched most of the verb semantic knowledge ultimately required. Our current axioms are "generic", but we will require much more specific ones. For example, the exact entailments of the verb *open* depend very much on what is being opened – a door, a book, a wine bottle, a mouth, a briefcase, a store, a festive event, etc. Our assumption is that capturing such entailments will require an "object-oriented" approach, i.e., one that draws upon the properties and methods associated with particular object types. We know the physical form, kinematics, dynamics, and function of doors, books, wine bottles, and so forth, and the use of a verb like *open* with a particular type of object seems to align the entailments of the verb with the known potentialities for that type of object. The work of James Pustejovsky and his collaborators on the Generative Lexicon (Pustejovsky 1991) is clearly relevant here. For example, he points out the different interpretations of the verb *use* in phrases such as *use the new knife (on the turkey)*, *use soft contact lenses*, *use unleaded gasoline (in a car)*, *use the subway*, etc., where these interpretations depend on the "telic qualia" of the artifacts referred to, i.e., their purpose and function.

5.2   Acquiring pattern-like world knowledge

A project we initiated more than 10 years ago is dubbed Knext (General **KN**owledge **EX**traction from **T**ext). Unlike KE efforts that glean specific facts about named entities from textual data – birth dates, occupations, geographic locations, company headquarters, product lines, etc. – our goal from the outset was the acquisition of simple *general* facts. The underlying idea was that sentences of miscellaneous texts, including realistic fiction, indicate common patterns of relationships

and events in the world, once inessential modifiers have been stripped away and specific entities have been generalized to entity types.

For example, consider the following sentence (from James Joyce's *Ulysses*):

Mr Bloom stood at the corner, his eyes wandering over the multicoloured hoardings.

Using the Charniak parse

```
(S1 (S (NP (NNP Mr) (NNP Bloom))
       (VP (VBD stood)
          (PP (IN at)
             (NP (DT the) (NN corner)))
          (, ,)
          (S (NP (PRP$ his) (NNS eyes))
             (VP (VBG wandering)
                (PP (IN over)
                   (NP (DT the) (JJ multicoloured) (NNS hoardings)))
                (. .)))))),
```

KNEXT applies compositional interpretive rules to obtain abstracted logical forms for several parts of the sentence that directly or implicitly provide propositional content. These propositional components are separately returned and automatically verbalized in English. The results are[9]

```
(:i <det man*.n> stand.v (:p at.p <the corner.n>))
A man may stand at a corner

(:i <the (plur eye.n)> wander.v (:p over.p <the (plur hoarding.n)>))
Eyes may wander over hoardings

(:i <det (plur eye.n)> pertain-to.v <det male*.n>)
Eyes may pertain-to a male

(:i <det (plur hoarding.n)> multicoloured.a)
Hoardings can be multicoloured
```

Note that in the first of the four factoids, *Mr Bloom* has been generalized to *a man*, and the modifying clause has been omitted. However, that clause itself gives rise to the second factoid, from which the possessive relation and the modifer *multicoloured* have been omitted. These modifiers in turn provide the remaining two factoids (where *pertain-to* is used as a noncommittal initial interpretation of the possessive). In the next subsection we mention some methods for partially disam-

---

[9]In preliminary logical forms, the ':i' keyword indicates an infix formula, and angle brackets indicate *unscoped* quantifiers. Determiner 'det' is used for an indefinite determiner; definites such as *the corner* are ultimately converted to indefinites as well, except in the case of ever-present "local entities" such as *the weather* or *the police*.

biguating factoids and introducing quantifiers. In particular, we will show "sharpened", quantified formulas that are derived automatically from the first and third factoids.[10]

In this way we typically obtain two or more general factoids per sentence in miscellaneous texts such as those comprising the Brown Corpus, The British National Corpus, Wikipedia, and weblog corpora. The first two of these sources provided several million factoids, and the latter two about 200 million (Gordon et al. 2010b). As some indication of the variety of factoids obtained, here is a small selection of some of the more interesting ones:

```
A bird may have feathers
A person may see with binoculars
Cats may meow
A cat may be chased by a dog
A baby may be vulnerable to infections
People may wish to be rid of a dictator
A male may refuse to believe a proposition
A female may tell a person to put_down a phone
A pen may be clipped to a top of a breast pocket
Factoids can be arcane
```

Generally above 80% of the factoids are rated by human judges as reasonable, potentially useful general claims. Among the publications on this work detailing the extraction methodology, filtering, refinement, and evaluation, the earliest was (Schubert 2002); a more recent one focusing on comparison with related work was (Van Durme and Schubert 2008).

As indicated at the beginning of the section, we have come to view such factoids as a separate but important knowledge category. One of the potential applications is in guiding a parser. For example, it is easy to see how the first two factoids in the above list could help select the correct attachments in the two versions of the sentence

He saw the bird with {binoculars, yellow tail feathers}.

In general, we believe that accurate parsing will require guidance by a multitude of syntactic and semantic patterns, where matches to a pattern "reinforce" particular combinatory choices (Schubert 1984, 2009).

---

[10]Concerning the apparent ambiguity of *wander* in the second factoid, a theoretical position one might take is that the meaning of this verb is actually the same whether the wandering is done by persons, eyes, or minds – but the *entailments* of this verb (and of verbal predicates more generally) depend on the types of the arguments. When persons wander, they physically move about; when eyes wander, it is the target of the gaze that shifts; when minds wander, it is the subject matter contemplated that shifts. An alternative to this position is that argument patterns, potentially in combination with other contextual cues, disambiguate the sense of the verb, and it is the disambiguated senses that carry the preceding entailments.

(This view seems in principle compatible with feature-based discriminative parsing models such as in (Huang 2008), though the features in the latter involve specific word and structural patterns, rather than any meaningful semantic patterns.) From this perspective, conformity with a particular abstract syntax is just one factor in the final analysis of a sentence; recognition of stock phrases, idioms, and patterns of predication and modifications are powerfully influential as well.

The second application of general factoids, which we have been actively exploring, is as a starting point for generating inference-enabling, quantified knowledge, as discussed in the following subsection.

5.3   Acquiring inference-enabling world knowledge

Statements such as that "*A bird may have feathers*" and "*Cats may meow*", though indicative of relationships and events apt to be encountered in the world, do not directly enable usefully strong inferences. For example, we cannot conclude that a particular bird named Tweety is likely to have feathers, only that this is a possibility. In order to obtain inference-enabling knowledge, we have sought to strengthen KNEXT factoids into quantified formulas in two different ways. One is aimed at inference of argument types, given certain relationships or events, while the other is aimed at inferring relationships or events, given certain argument types. A simple example of these alternative ways of strengthening a factoid would be to strengthen "*A bird may have feathers*" into

> *If something has feathers, it is most likely a bird,*

and

> *All (or at least most) birds have feathers.*

Our methodology for the first type of strengthening involves collecting argument types for a given verbal predicate, and seeking a small set of Wordnet hypernyms that cover those types. In the case of the factoid above we would collect factoids matching the pattern "*A(n) ?P may have feathers*", and then look for a WordNet hypernym that covers (some senses of) the nominals matching *?P* (or at least a majority of them). In the present case, we would obtain (the primary senses of) *bird* and *person* as a covering set, and hence derive the quantified claim that "*most things that have feathers are either birds or persons.* (Persons unfortunately cannot easily be eliminated because of such locutions as "ruffled his feathers", and also because of default assumptions that, e.g., *their feathers* refers to human possessors.)

This method turned out to be quite effective (Van Durme et al. 2009), and – serendipidously – provided a method of disambiguating

the senses of nominal argument types. For example, the capitalized types in the following factoids were disambiguated to their literary communication senses, thanks to their subsumption under a common *communication* hypernym,

> *A child may write a* LETTER
> *A journalist may write an* ARTICLE,

thereby excluding such senses as alphabetic letters, varsity letters, grammatical articles, and articles of merchandise. However, while we have demonstrated the efficacy of the method on limited samples, we have not yet applied it to the full KNEXT factoid base. The problem is that the requisite WordNet searches for a small (not too general, not too specific) set of covering hypernyms can be very expensive. For example, the search for a set of hypernyms covering a majority of types matching *?P* in "*A person may get a(n) ?P*" would be quite time-consuming, in view of the many complement types allowed by *get*: *chance, babysitter, doctor, checkup, pizza, newspaper, break, idea, job, feeling, point, message, impression, kick, look, ticket, etc.*

We refer to the second type of strengthening, aimed at finding probable relationships and action/event types for given types of individuals, as *factoid sharpening* (Gordon and Schubert 2010). This method is not dependent on abstraction from groups of factoids, but rather strengthens individual factoids (except that we favor factoids that have been abstracted several times from different sources). This may at first seem suspect: A factoid says what it says – how can it be coerced into saying more? The answer lies in the semantic categories of the predicates involved. For example, a factoid like "*A bird may have feathers*" instantiates the predicate pattern '*A(n)* ANIMAL *may have* ANIMAL-PART". Since we know that part-of relationships are typically uniform across living species, and permanent, we can fairly safely sharpen the factoid to a quantified fomula asserting that *all or most birds permanently have some feathers as a part*. Note that just like the first method of strengthening factoids, the sharpening process can also lead to predicate disambiguation; in the present example, uncommon senses of *bird* and *feather* are eliminated, and *have* is specialized to *have-as-part*.

The following are the sharpened formulas automatically derived from two of the factoids displayed earlier (obtained from the James Joyce sentence). The expression `[x | e]` in the last formula denotes the ordered pair (or list) consisting of the semantic values of `x` and `e`; such agent-episode pairs represent specific, temporally located actions or attributes of an agent in EL.

```
(:i <det man*.n> stand.v (:p at.p <the corner.n>))
```

```
A man may stand at a corner

(:i <det (plur eye.n)> pertain-to.v <det male*.n>)
Eyes may pertain-to a male

  (many x: [x man.n]
     (occasional e
        (some y: [y corner.n]
             [[x stand-at.v y] ** e])))

  Many men occasionally stand at a corner.

  (all-or-most x: [x male.n]
     (some e: [[x | e] enduring]
        (some y: [y eye.n]
             [[x have-as-part.v y] ** e])))

  All or most males have an eye as a part.
```

We take such formulas as justifying inferences such as that if John is a man, then it is rather likely that he occasionally stands at some corner, and very likely that he has an eye as a part. (The probabilistic qualifiers are presumed to apply in the absence of other knowledge bearing on the conclusions.)

Of course, obtaining such results requires rules that match various patterns of predication against factoids, and generate appropriate sharpened formulas corresponding to successful matches. This has been facilitated by two broadly useful devices: predicate classification with the aid of programs that make use of WordNet relations and VerbNet classes; and a template-to-template transduction system called TTT (Purtee and Schubert 2012). Some subject and object types that are important in transducing factoids into sharpened formulas are ones with WordNet hypernyms *causal agent, body part, professional, food, artifact, event, state, psychological feature*, and several others. Properties of sentential predicates that are important include repeatability (contrast *swim, marry*, and *die*), stativity (contrast *swim, believe*, and *lawyer*), and ones headed by special verbs such as copular *be* and relational *have*. Small sets of VerbNet classes and WordNet relations often enable identification of the relevant properties.

We have produced more than 6 million sharpened formulas, and evaluated the quality of samples by human judgement. When the factoids supplied as inputs to the sharpening process were restricted to ones assessed as being of good quality, about 55.4% of the resulting formulas were judged to be of good quality (for details see Gordon and Schubert 2010); unsurprisingly, for unscreened factoids the percentage of good sharpened formulas was lower, 36.8%. Though the results are

encouraging, clearly much remains to be done both in screening or improving input factoids and doing so for resultant sharpened formulas. One possible approach is use of Amazon's Mechanical Turk (see Gordon et al. 2010a). For some further refinements of the sharpening work, aimed at refining event frequencies in general claims (e.g.,"*If a person drives taxis regularly, he or she is apt to do so daily or multiple times a week*"), see (Gordon and Schubert 2012).

Finally, we have also begun to extract "if-then" knowledge from text based on discourse cues, in particular, on locutions that indicate failed expectations, expected or hoped-for results, or "good-bad" contrasts (Gordon and Schubert 2011). For example, from the sentence

> *The ship weighed anchor and ran out her big guns,*
> *but did not fire a shot,*

it can be plausibly conjectured that

> *If a ship weighs anchor and runs out her big guns,*
> *then it may fire a shot.*

However, this work to date generates hypotheses in English only; logical formulation of these hypotheses is left to future work.

6   Can we acquire knowledge by direct interpretation of general statements?

None of the above approaches to acquiring lexical and world knowledge are dependent on genuine understanding of even a limited range of ordinary language. This fact testifies to the elusiveness of a long-standing AI dream: to develop a NLU system stocked with just enough linguistic and world knowledge to be able to learn from lexical glosses, encyclopedias and other textual knowledge repositories, and in this way achieve human-like, or even trans-human, competence in language understanding and commonsense reasoning.

We will not attempt here to survey work to date towards achieving this dream. Rather, we will outline some of the challenges confronting such an enterprise, from an Episodic Logic/EPILOG perspective. After all, EL is close in its form and semantic types to ordinary language, so that transduction from surface language to an EL meaning representations ought to be relatively easy; easier, for instance, than mapping English to CycL (Lenat 1995), whose representations of linguistic content, though logically framed, bear little resemblance to the source text.[11]

---

[11]Without belittling this important effort, we should add that the Cyc project has not yet gathered nearly enough knowledge for general language understanding and commonsense reasoning; based on our perusal of KNEXT factoids, and on knowledge

Computing initial logical forms in Episodic Logic is in fact fairly straightforward – a refinement of the compositional methods used in Knext. Some issues we should note before moving on to the more formidable challenges in fully interpreting sentences intended to express general knowledge are the following:

- word sense ambiguity (e.g., *crown* in the WordNet gloss for *tree*, which is not disambiguated in the "glosstag" data provided for WordNet (`http://wordnet.princeton.edu/glosstag.shtml`);
- distinguishing between verb complements and adjuncts, such as the two PPs in *communicating with someone with hand signals*;
- identifying the syntactic/semantic role of "SBAR" constituents, which can be relative clauses, clausal adverbials, or clausal nominals; for example, note the ambiguity of "*There are many villages where there is no source of clean water*";
- attachment decisions, for PPs and other phrases (e.g., The Open Mind sentence "*Sometimes writing causes your hand to cramp up*" is parsed by our off-the-shelf parser with a subject *sometimes writing*, viewed as an adverb modifying a progressive participle);
- coordinator scope (e.g., the WordNet gloss for *desk* is "*a piece of furniture with a writing surface and usually drawers or other compartments*", which our parser mangles rather badly, forming a conjunction "*with a writing surface* **and** *usually*" and a disjunction "*furniture … **or** other compartments*");
- ellipsis (this is more important in discourse than in statements of general knowledge);
- anaphora (e.g., from Simple Wikipedia, "*The roots of a tree are usually under the ground. One case for which* **this** *is not true are the roots of the mangrove tree*"; as noted below, such complexities contraindicate reliance on this source).

Assuming that many of these problems can be overcome using syntactic and semantic pattern knowledge, can we then obtain inference-enabling knowledge of the sort needed for language understanding and commonsense reasoning from the sources we have mentioned?

There are two sorts of challenges: challenges inherent in the kinds of knowledge provided by the targeted sources and the way the knowledge is presented, and challenges in interpreting generic sentences. The primary sources we are considering are WordNet glosses, Simple Wikipedia entries, and Open Mind factoids.

---

gaps observed when posing questions to ResearchCyc, we conjecture that at least tens of millions of individual knowledge items will be required.

6.1 Problems with the sources

One problem with WordNet is that glosses for verbs resolutely refrain from mentioning verb objects, let alone their types. For example, the glosses for the verbs *saw* and *lisp* are *cut with a saw* and *speak with a lisp*, respectively, leaving any automated system guessing as to whether either of these verbs requires an object, and if so, what kind. Consultation of the examples, such as "*saw wood for the fireplace*" can sometimes resolve the ambiguity, but there may be no examples (as in the case of *lisp*), or parsing the examples may itself be problematic.

The descriptions of physical objects in WordNet provides very few constraints on the typical structure and appearance of the objects. For example, the description of a tree as (in part) "*a tall perennial woody plant having a main trunk and branches ...*" does not distinguish the height of a tree from other tall plants, such as sunflowers, or provide a clear picture of the verticality of the trunk and configuration of branches relative to it.[12] It seems that a few internal geometric prototypes (representing, say, a deciduous tree, an evergreen and a palm), traversable by property abstraction algorithms, would be worth a thousand words in this regard. (For a recent paper on the computational learnability of geometric prototypes, and their utility in vision and tactile sensing, see (Yildirim and Jacobs in press).)

A more serious problem from an NLU and commonsense reasoning perspective is that the gloss is silent on the roles trees play in our world and in our stories – providing shade, shelter, fruits, a pleasing sight, a climbing opportunity, a habitat for birds and other creatures, wood for buildings, etc. Can these gaps be filled in from sources like Simple Wikipedia? The initial 240-word summary in the entry for *tree* unfortunately contains few of the generalities we are looking for, while the 2700-word entry taken in its entirety touches on more of them (though still not mentioning shade, fruits, birds, or building materials), but these are scattered amidst many less relevant items, such as the aerial roots of the Banyan tree, the xylem and phloem cells comprising wood, the significance of growth rings, and so on. Consequently, gaining basic knowledge from such entries already presupposes a far more sophisticated language understanding system than we can currently envisage.

The Open Mind Common Sense project at MIT (Singh 2002; Singh et al. 2002) comes much closer to providing the kinds of knowledge we are seeking (e.g., the various roles of trees mentioned above), cast in the form of simple, separately interpretable statements. The kinds of

---

[12]The definition as a *tall* plant does not directly entail a vertical trunk; consider a *windjammer* described as a *tall sailing ship having a steel hull.*

statements obtained are constrained by the way information is solicited from contributors. Consequently, many of the more subtle kinds of factoids obtained by Knext are unlikely to show up in Open Mind (e.g., the earlier example, "*A baby may be vulnerable to infections*", or "*A pen may be clipped to the top of a breast pocket*"). But on the other hand, the Open Mind factoids (exclusive of Verbosity-derived ones) seem more reliable at this point, as we have not yet used filtering of Knext factoids by crowd-sourcing on a large scale. Also, Open Mind factoids often include desires, propensities, uses, and locales that are much rarer in Knext, such as "*A baby likes to suck on a pacifier*", "*Pens can be used to write words*", or "*You are likely to find a pen in an office supply store*".

However, many of the statements are awkardly phrased as a result of the particular 20 questions posed to contributors, and for our interpretive purposes would require either a specially adapted parser or preliminary rule-based transformations into more natural English. For example,

> *Somewhere a bird can be is on a tree*

would be more readily interpretable as

> *A bird can be on a tree.*

(Jonathan Gordon at the University of Rochester has implemented a few such transformations based on the previously mentioned tree-to-tree transduction tool TTT.) Also, many of the statements (including the above) share a weakness with our Knext factoids in that they express mere possibilities, providing little support for inference. We would expect to apply our sharpening methods in many such cases.

In addition, some of the Open Mind statements reflect the hazards of crowdsourcing. Integration of Verbosity data into Open Mind, in particular, seems to have contributed a fair share of fractured-English examples, such as "*Coat is looks lot*", "*Chicken is a saying*", and "*Car is broom-broom*". Also, some statements blend myth and reality. An example is "*Something you find under a tree is a troll*", which is annotated with the same degree of support as, for instance, "*A tree can produce fruit*". Fortunately, however, Open Mind also contains the statement that "*A troll is a mythical creature*".

Perhaps the most serious problem (also shared with many Knext factoids) is the radically underspecified nature of many claims, such as that "*Cars can kill people*", "*Fire can kill*", "*A knife is used for butter*", or "*Milk is used for cream*"; these give no indication of the agent, patient, or event structure involved, and as such could be quite misleading. While we can imagine filtering out many such terse state-

ments in favor of more elaborate ones, such as "*A knife is used for spreading butter*" and "*Knives can butter bread*" (also found in Open Mind), ultimately we need to confront the fact that generic statements are almost invariably underspecified in various ways. This relates to the observations in the following subsection.

6.2   Problems with interpretation of generics

There are subtle interpretive issues, familiar in the literature on generic sentences, that would need to be addressed (e.g., Carlson and Pelletier 1995). For example, "*A tree can grow*" should not be interpreted as referring to a particular tree, but rather to trees in general. Contrast the sentence with "*An asteroid can hit the Earth*", which is unlikely to be understood as a generalization about asteroids. Moreover, while we would regard "*A tree can grow*" as essentially equivalent to "*Trees can grow*", where on G. Carlson's analysis, *trees* denotes the *kind, trees*, the singular indefinite subject in the first version cannot be so viewed, as is clear from the contrast between "*Trees are widespread*" and #"*A tree is widespread*".

A general approach to computing the logical forms of generic (as well as habitual) sentences, covering the above phenomena and others, appears to require the introduction of various syntactically null constituents, including kind-forming operators, quantifying adverbials, and presupposed material, all under various soft constraints dependent on such features as telicity and the sortal categories of predicates (e.g., individual-level vs. stage-level, and object-level vs. kind-level). For example, consider the generic sentence

Dogs bark.

As a preliminary logical form, abiding by Carlson's analysis of bare plurals, we would obtain

```
[(K (plur dog)) bark].
```

However, this is sortally inconsistent, since *bark* is a predicate applicable to objects, rather than kinds (in contrast with predicates like *evolve* or *are widespread*). Thus we would elaborate the LF to

```
[(K (plur dog)) (generally bark)],
```

where *generally P* is understood as something like "*is a kind such that all or most of its realizations have property P*". In other words, the quantifying adverbial (Q-adverbial) *generally* in effect type-shifts[13] an object-level predicate to a kind-level predicate. However, this is still sortally faulty, because *bark* is not an individual-level predicate (one

---

[13]More accurately, sortally shifts.

simply true or false of an individual viewed as space-time entity spanning its entire existence), but rather an episodic predicate (true or false of an individual *at* a particular temporally bounded episode). Therefore, a further operator is called for, one that converts an episodic predicate into an individual-level predicate. A natural choice here seems to be a *habitual* operator expressing *(at least) occasionally*:

```
[(K (plur dog)) (generally (occasionally bark))]
```

This corresponds to the quite natural reading of the original sentence, expressing that dogs *in general* bark *(at least) occasionally*. EL and Epilog allow us to work directly with such logical forms, but we can also easily use them, in conjunction with meaning postulates about *K, generally*, and *occasionally* to derive representations closer to FOL-like ones (though with loss of the law-like content of the original version):

```
(all-or-most x: [x dog]
  (exist-occasional e [[x bark] ** e])).
```

There is a plausible alternative to the operator *occasionally*, namely *can*, expressing an ability or propensity; this also converts an episodic predicate into an individual-level predicate. After conversion to the explicit quantified form, we would have

```
(all-or-most x: [x dog] [x (can bark)])
```

We note that the Open Mind instructions requesting input from contributors preclude simple habitual generics such as "*Dogs bark*", favoring ability-generics instead. Indeed the highest-rated entry for *dog* is

　　An activity a dog can do is bark,

whose phrasing we would simplify to "*A dog can bark*".

　　Without going into additional details, we note some further issues that arise in the interpretation of generics, as illustrated by the following sentences.

　　Dogs are occasionally vicious
　　Dogs can mother puppies
　　Dogs often chase mailmen

Note that the first sentence is ambiguous between a reading that quantifies over dogs ("*The occasional dog is vicious*") and one that ascribes a habitual behavior (occasional episodes of viciousness) to dogs in general. These alternatives seem to arise from a basic atemporal/temporal ambiguity that Q-adverbs are subject to; in the present case, *occasionally* can mean either "*is a kind some of whose realizations …*", or "*at some times …*", where in the latter case we would also assume an implicit *generally*, i.e., the sentence expresses that "*Dogs in general are*

at some times vicious". Which reading comes to the fore in a given generic sentence depends on whether the sentential predicate is itself individual-level or episodic, or ambiguous between these, as in the case of *vicious*. The second sentence (from Open Mind) illustrates the fact that a proper understanding of a generic sentence often requires imposition of constraints on the kind in subject position, based on common knowledge; i.e., here we need to restrict the kind under consideration to *female* dogs. The third sentence (also from Open Mind) illustrates that in addition, we may need to bring into play presuppositions based on world knowledge. We know that mailmen are not perpetually available for chasing by a given dog, but only when delivering mail in the immediate vicinity of the dog's abode. So in effect we understand the sentence as saying something like

> *When a mailman delivers mail in the vicinity of a dog's abode, the dog often chases the mailman.*

Arriving at such interpretations automatically seems further out of reach than in the case of the earlier examples – we would have to use general facts about mailmen, and about mail delivery to homes, and about the presence of dogs in and around homes, to reach the desired interpretation of the sentence in question.

7 Concluding remarks

We showed that the representational and inferential style of Episodic Logic and the Epilog system are close to those of Natural Logic. But we also gave ample evidence showing that our approach and system allow for a broader range of inferences, dependent on world knowledge as well as lexical knowledge. Moreover, Epilog performs goal-directed inference without being told from what specific premises to draw its conclusion (as in recognizing textual entailment), and can also perform some forward inferences.

It is sometimes assumed that Natural Logic has an advantage over symbolic logic in that it is tolerant of ambiguity, vagueness, and indexicality. However, Epilog also tolerates ambiguous, vague or indexical input, as was seen in the examples concerning the availability of a *small crane* for rubble removal, and the location of *most of the heavy equipment* in a part of Monroe county. At the same time, we need to recognize that there are limits to how much vagueness of ambiguity can be tolerated in a knowledge-based system, lest it be led astray. For example, a system with common sense should not conclude from "*Bob had gerbils as a child*" that Bob consumed, or gave birth to, small rodents as a child (as a method based on alignment, polarity and word-level editing

might well conclude).

We argued that the greatest challenge in achieving deeper understanding and reasoning in machines is (still) the knowledge acquisition bottleneck, and outlined our multiple lines of attack, all with some degree of success, on lexical and world knowledge acquisition. The methods for lexical knowledge acquisition have relied on sources such as distributional similarity data, WordNet hierarchies, VerbNet classes, and collections of implicative and other verbs that can easily be mapped to axioms supporting NLog-like inferences. Our approach to general knowledge extraction from text has delivered many millions of simple general factoids, which we suggested were potentially useful for guiding parsers, and which we showed to be capable of yielding inference-enabling quantified formulas through techniques based on groups of factoids with shared verbal predicates, or based on *sharpening* of individual factoids with the help of additional knowledge sources.

But the knowledge procurable in these ways provides only a fraction of the kinds of general knowledge people appear to employ in language understanding and reasoning, and we are therefore committed in future work to continuing the development of interpretive methods that can produce generally accurate, logically defensible, inference-enabling interpretations of verbally expressed general knowledge in such sources as WordNet glosses and Open Mind. The challenges that we outlined in considering such an enterprise are certainly formidable but seem to us ones that can be met, using the NLP resources already at hand along with pattern-based parser guidance, and methods for interpreting generic sentences grounded in linguistic semantics.

Acknowledgments

References

Carlson, G.N. 1977. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1(3):413–456.

Carlson, G.N. and F.J. Pelletier, eds. 1995. *The Generic Book*. Chicago and London: Univ. of Chicago Press.

Clausen, David and Christopher D. Manning. 2009. Presupposed content and entailments in natural language inference. In *ACL-IJCNLP Workshop on Applied Textual Inference*.

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

Fadiga, L., L. Craighero, M. Fabbri Destro, L. Finos, N. Cotilon-Williams, A.T. Smith, and U. Castiello. 2006. Language in shadow. *Social Neuroscience* 1(2):77–89.

Fedorenko, E., J. Duncan, and N. Kanwisher. 2012. Language-selective and domain-general regions lie side-by-side within Broca's area. *Current Biology* 22:1–4.

Gordon, J. and L.K. Schubert. 2012. Using textual patterns to learn expected event frequencies. In *NAACL-HLT Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX)*. Montreal, Canada.

Gordon, Jonathan and Lenhart K. Schubert. 2010. Quantificational sharpening of commonsense knowledge. In *Proc. of the AAAI 2010 Fall Symposium on Commonsense Knowledge*.

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

Gordon, Jonathan, Benjamin Van Durme, and Lenhart K. Schubert. 2010a. Evaluation of commonsense knowledge with Mechanical Turk. In *Proc. of the NAACL Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Los Angeles, CA.

Gordon, Jonathan, Benjamin Van Durme, and Lenhart K. Schubert. 2010b. Learning from the Web: Extracting general world knowledge from noisy text. In *Proc. of the AAAI 2010 Workshop on Collaboratively-built Knowledge Sources and Artificial Intelligence (WikiAI)*. Atlanta, GA.

Huang, Liang. 2008. Forest reranking: Discriminative parsing with non-local features. In *Proc. of ACL-08: HLT*, pages 586–594. Columbus, OH.

Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4:167–193.

Karttunen, Lauri. 2012. Simple and phrasal implicatives. In *Proc. of *SEM: The First Joint Conf. on Lexical and Computational Semantics*, pages 124–131. Montréal, Canada.

Lenat, Doug. 1995. CYC: A large-scale investment in knowledge infrastructure. *Comm. of the ACM* 38(11):33–38.

MacCartney, Bill and Christopher D. Manning. 2008. Modeling semantic containment and exclusion in natural language inference. In *Proc. of the*

22nd Int. Conf. on Computational Linguistics (COLING '08), pages 521–528.

MacCartney, Bill and Christopher D. Manning. 2009. An extended model of natural logic. In *Proc. of IWCS-8*.

Morbini, Fabrizio and Lenhart K. Schubert. 2009. Evaluation of Epilog: a reasoner for Episodic Logic. In *Commonsense 09*. Toronto, Canada.

Nairn, R., C. Condoravdi, and L. Karttunen. 2006. Computing relative polarity for textual inference. In *Inference in Computational Semantics (ICoS-5)*, pages 67–76.

Purtee, A. and L.K. Schubert. 2012. TTT: A tree transduction language for syntactic and semantic processing. In *EACL 2012 Workshop on Applications of Tree Automata Techniques in Natural Language Processing (ATANLP 2012)*. Avignon, France.

Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics* 17(4):409–441.

Schubert, L.K. 2009. Language understanding as recognition and transduction of numerous overlaid patterns. In *AAAI Spring Symposium on Learning by Reading and Learning to Read*, pages 94–96. Stanford, CA.

Schubert, L.K., J. Gordon, K. Stratos, and A. Rubinoff. 2011. Towards adequate knowledge and natural inference. In *AAAI Fall Symposium on Advances in Cognitive Systems*. Arlington, VA.

Schubert, Lenhart K. 1984. On parsing preferences. In *Proc. of the 10th Int. Conf. On Computational Linguistics (COLING-84)*, pages 247–250. Stanford Univ., Stanford, CA.

Schubert, Lenhart K. 2000. The situations we talk about. In J. Minker, ed., *Logic-Based Artificial Intelligence*, pages 407–439. Kluwer, Dortrecht.

Schubert, Lenhart K. 2002. Can we derive general world knowledge from texts? In *Proc. of HLT02*.

Schubert, Lenhart K. and Chung Hee Hwang. 2000. Episodic Logic Meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In L. Iwanska and S. Shapiro, eds., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*. Menlo Park, CA, and Cambridge, MA: MIT/AAAI Press.

Schubert, Lenhart K., Benjamin Van Durme, and Marzieh Bazrafshan. 2010. Entailment inference in a natural logic–like general reasoner. In *Proc. of the AAAI 2010 Fall Symposium on Commonsense Knowledge*.

Singh, Push. 2002. The public acquisition of commonsense knowledge. In *Proc. of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*.

Singh, P., T. Lin, E.T. Mueller, G. Lim, T. Perkins, and W.L. Zhu. 2002. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proc. of the Confederated Int. Conf. DOA, CoopIS and ODBASE*, pages 1223–1237. Irvine, CA.

Stent, Amanda. 2000. The Monroe Corpus, Tech. Rep. no. TR728 and TN99-2. Tech. rep., Dept. of Computer Science, Univ. of Rochester, Rochester, NY, USA.

Stratos, Karl, Lenhart Schubert, and Jonathan Gordon. 2011. Episodic Logic: Natural logic + reasoning. In *Int. Conf. on Knowledge Engineering and Ontology Development (KEOD)*. Paris, France. Available (with INSTIC/Primoris login) at http://www.scitepress.org/DigitalLibrary.

Valencia, V. Sánchez. 1991. *Studies on Natural Logic and Categorial Grammar*. Ph.D. thesis, University of Amsterdam.

van Benthem, J. 1991. *Language in Action: categories, lambdas and dynamic logic*, vol. 130. Amsterdam: Elsevier.

van Benthem, J. 2007. A brief history of natural logic. In *Int. Conf. on Logic, Navya-Nyāya & Applications: Homage to Bimal Krishna Matilal*.

Van Durme, Benjamin, Phillip Michalak, and Lenhart K. Schubert. 2009. Deriving generalized knowledge from corpora using WordNet abstraction. In *12th Conf. of the Eur. Chapter of the Assoc. for Computational Linguistics (EACL09)*. Athens, Greece.

Van Durme, B. and L.K. Schubert. 2008. Open knowledge extraction using compositional language processing. In *Symposium on Semantics in Systems for Text Processing (STEP 2008)*. Venice, Italy.

van Eijck, J. 2005. Natural Logic for Natural Language. `http://homepages.cwi.nl/\verb+~+jve/papers/05/nlnl/NLNL.pdf`.

Yildirim, I. and R.A. Jacobs. in press. Transfer of object category knowledge across visual and haptic modalities: 3 experimental and computational studies. *Cognition* .