

What Kinds of Knowledge are Needed for Genuine Understanding?*

Lenhart Schubert
University of Rochester
schubert@cs.rochester.edu

Abstract

Crucial considerations in launching knowledge acquisitions projects aimed at human-like understanding and reasoning include the choice of semantic/knowledge representation (SR/KR), the types of knowledge required and their use in understanding and inference, and approaches to acquiring that knowledge. This paper argues that the SR/KR needs to resemble language in its expressivity and needs to support complex reasoning, and outlines some of the most important knowledge requirements, and some steps towards acquiring that knowledge.

1 Introduction: NLU in historical perspective

In referring to *kinds* of knowledge, I'm referring to both the *form* and the *content* of knowledge. By *genuine* understanding I mean the kind of understanding that enables reasoning with the content derived from language, in concert with background knowledge; I should add that my interest is in broad understanding rather than narrow task-oriented understanding where the a priori constraints on content allow the use of heuristic rules for extracting expected types of data.

Understanding was the goal of much of the most prominent natural language processing research from the late 1960s to the middle 80s; and exciting progress was indeed made, enabling inferential question answering for brief stories of various types, including simple restaurant tales, fairy tales and diplomatic visit reports. The chief insight was that tremendous amounts of knowledge are needed for understanding and inference, and in particular that much of the required knowledge is in the form of stereotyped schemas or "scripts". For example, there was Roger Schank's well-known "restaurant script", which enabled comprehension of simple stories such as the following:

John had lunch at Mario's. The hamburger he ordered was tasteless, so he left a small tip.

The script and a variety of high-level entailment rules allowed answering questions whose answers are not explicitly given

*This work was supported in part by DOD/NAVY/ONR Award # N00014-11-1-0417.

in the story, such as "was John given a hamburger?", "Did he eat it?", "Where?", and so on.

So what brought these seemingly promising developments to a virtual halt over the following decades? Basically, the problem was lack of scalability. Several factors contributed to this limitation. First of all, parsing was quite inaccurate. For example, there was (and remains) the infamous prepositional phrase (PP) attachment problem, as in

John observed a bird {with binoculars, with yellow tail feathers, with delight, with its brood, with his fellow birders, etc.}.

In that sentence the PP can be attached so as to modify either the verb *observed* or the object, *bird*; in particular in the case of *binoculars*, *delight*, and *fellow birders* we naturally associate the PP with the act of observation, whereas for *yellow tail feathers* and *its brood of chicks* we perceive the PP unambiguously as modifying *bird*.

An even greater problem, also persisting to this day, was the knowledge acquisition bottleneck. How could we acquire the myriad knowledge items needed for understanding? These include a broad spectrum of ontological knowledge, numerous well-known facts such as the categories of named entities that occur frequently in text, and most of all various kinds of general world knowledge at both the lexical level, such as that restaurants serve meals to customers for pay, and at the more elaborate stereotyped story level, such as is required for the simple story just illustrated.

Certain limitations that were less widely recognized were inherent in many of the semantic representations and knowledge representation used, in particular their inadequate expressivity, unanalyzed semantic type structure, and unnecessary "distance" from language. Without adequate expressivity, information is lost or distorted. Without an understanding of semantic type structure, reasoning methods are apt to derive nonsense. And if the representation is remote from language, the interplay between knowledge and language becomes difficult to implement.

2 The ML revolution

The 80s and 90s of course saw the beginning and increasing dominance of the machine learning (ML) revolution, made possible by new developments in learning theory, both statistically based and neural-net based, and by the burgeoning

quantities of text that became available on the internet. Both approaches exploit the distributional properties of language in large data sets. For example, in the familiar sentence

Time flies like an arrow,

which has 4 or 5 possible parses and corresponding interpretations, it is arguably the co-occurrence frequency of *time* and *flies* as sentential subject and main verb respectively that favors the intuitively natural interpretation, whereas *time flies* as a noun-noun combination analogous to *fruit flies* or as an imperative verb plus object is virtually unheard of in discourse.

The ML community can certainly point to many noteworthy and important successes, particularly in areas where relatively shallow methods suffice and where a certain level of error (from, say, 5% to 40%, depending on the area of application and how you measure error rates) are acceptable. These applications include greatly improved speech recognition, learning of probabilistic grammars, machine translation, document retrieval, factoid and rule learning from textual data, and so forth. Research towards the goal of genuine, broad understanding and commonsense reasoning benefited peripherally from these advances, yet achievement of the goals remains well out of reach.

The key challenges that remain, as argued below, are (a) designing a semantic representation/knowledge representation (SR/KR) that is commensurate with natural language, (b) acquisition of the requisite tens of millions, perhaps even hundreds of millions, of pattern-like, schema-like and axiom-like knowledge items, and (c) effective, scalable soft and firm inference methods.

3 Representational requirements for understanding and reasoning

We cannot depend on machine learning to yield an SR/KR commensurate with natural language for various reasons. The most important is that machine learning cannot *invent* a meaningful representational framework; it can only learn what you provide examples of or build into latent variables:

- if you annotate data with class labels, ML can produce class labels;
- if you annotate sentences with parse trees, ML can produce parse trees;
- if you annotate sentences with translations, ML can produce translations (and latent syntax transduction rules, if we provide a syntactic framework for these);
- if you annotate English db queries with answers, ML can produce answers via formal queries (if the syntax of those queries has been supplied);
- if you annotate sentences with “abstract meaning representations”, ML can produce AMRs; etc.

We also cannot depend on ML to lead directly to human-like reasoning: Reasoning requires knowledge in an interpretable, modular form, whose knowledge items or modules can be deployed in a broad spectrum of topic areas, in miscellaneous combinations with one another. The notion propounded by some prominent researchers that ML has made

symbolic language understanding and reasoning methods obsolete should be dispelled by considering simple examples such as the following:

When Randy learned that his cancer was terminal, he decided to stop receiving chemotherapy and enrol in a hospice program designed to provide palliative care.

Why did Randy decide this?

I don't believe that this lies within the near-term reach of any current NLU system (without resort to targeted knowledge precoding), nor do I think that any current ML techniques, including “deep learning” can begin to crack it. Let's first consider what we, as readers, can infer about Randy's situation. After all, changing one small word, *learned*, to another, such as *dreamed*, would greatly alter our understanding of the actual situation. At least the following knowledge items seem essential:

- When you learn something, you then know it; (so Randy knew he had terminal cancer);
- When someone knows that something is the case, it is indeed the case; (so Randy had terminal cancer).
- Patients with a terminal illness usually die within a few months; (thus Randy was destined to die within a few months);
- Chemotherapy is intended to combat cancer and prolong life, but results in suffering and weakness;
- No medical treatment will significantly prolong a terminal cancer patient's life;
- Chemotherapy is a medical treatment;
- Stopping something entails (presupposes) that it was going on; (so Randy had been receiving chemotherapy, and therefore had endured suffering);
- Receiving palliative care reduces pain and discomfort;
- Enrolling in a program to provide certain services will lead to receiving those services; (so Randy would receive palliative care);
- If Randy continued receiving chemotherapy, he would endure further suffering without significant life extension, while palliative care would make him feel better;
- People act to optimize expected rewards and minimize suffering.

So if Randy knew all of the above, his choice of palliative care instead of chemotherapy is explained. But how did he know all this?

- Commonsense inferences that “jump out” at us also “jump out” at others (who possess the same premise knowledge): *simulative inference* (see a formalization in Kaplan & Schubert 2000).
- All the items above are common knowledge (especially to cancer patients), so the indicated inferences were just as obvious to Randy as they are to us.

Thus Randy knew about the likely consequences of continued chemotherapy vs. those of palliative care, and so under the presumption of optimization of expected outcomes,

Randy's choice is explained. (Again, we're also presupposing a lot of lexical and paraphrase knowledge: *terminal illness, dying, medical treatment, palliative care, optimize, life extension, suffering, etc.*)

If we are to capture such knowledge, we need to at least be able to express it internally in an inference-enabling form. Here are some noteworthy expressive devices encountered in the example:

- Predication, of course (Randy learned something, he received chemotherapy, and so on)
- Temporally related events (getting cancer, receiving chemotherapy, etc., and their implicit consequences)
- Causal relations (chemotherapy causes suffering, possible life prolongation)
- Conjunction and disjunction (Randy could choose continued chemotherapy or palliative care)
- Negation (No treatment can save a terminal patient's life)
- Quantification (most patients with metastatic cancer receive chemotherapy; almost everyone knows this)

These devices already suggest the need for at least the expressive power of first-order logic (FOL), indeed somewhat more, since the above quantifiers are nonclassical. And there are still more demands on expressivity:

- Genericity – almost all the general knowledge listed has the character of applying in “typical” cases, but allowing for exceptions;
- Patterns of events and behavior (the potential courses of events if you get cancer; patterns of interaction between doctors and patients); descriptions of such patterns can be thought of as *generic passages* (Carlson & Spejewski 1997);
- Modal/mental notions (*learning, knowing, expecting, deciding, inferring, intending, ...*)
- Counterfactuals (deciding against something because of the adverse consequences it *would* have);
- Uncertainty (Randy's life *expectancy*);
- Predicate modification (*terminal cancer, palliative care, continuing/stopping receiving chemotherapy, feel better*);
- Sentence modification (*probably, palliative care will ease Randy's suffering*);
- Predicate reification (receiving chemotherapy, suffering, palliative care)
- Sentence reification (Randy learned *that his cancer was terminal*).

Are any of these devices peculiar to English? Of course not – they are available in all languages. Some reasons for viewing language as a mirror of mind, i.e., for assuming that the expressive devices we find in human languages reflect innate, language-like semantic types—a “mentalese” cognitive framework—are the following.¹

¹Note that innate “mentalese” semantic types need not entail any *linguistic universals* in the grammatical sense. For example, there

- Full language and thought are conjectured to have arisen concurrently and rather recently—perhaps some 300,000-50,000 years ago (e.g., Ian Tattersall, Derek Bickerton, Philip Lieberman, and many others);
- Humans occupy the “cognitive niche” (Pinker) because they can learn, store, manipulate for inference, and communicate symbolically encoded thoughts – semantic/knowledge representations.
- Richard Montague showed that there is a tight, highly systematic relationship between linguistic structure and meaning (compositional semantics).
- The simplest explanation for how it is possible for language understanding and knowledge-based inference to function synergistically is that our internal SR/KR is itself language-like.

Many sorts of SR/KR have been suggested over the years, and all are in certain respects language-like. However, many are either expressively inadequate in terms of the semantic categories mentioned above, or inferentially inadequate, because of restricted expressivity or lack of a theory of semantic types that could guide formulation of inference rules. Since I attempted a fairly comprehensive summary and critique in (Schubert 2015), I won't repeat myself here, except to point to the enumeration of approaches below. The criticism indicated in the listing should be viewed as “caricatures” rather than unqualified assessments.

FOL, DRT (e.g., Allen, Jurafsky & Martin, Kamp, Heim, Bos, ...); [expressively inadequate]

Semantic nets (wide spectrum: Shapiro, Sowa, ConceptNet, ...); [any notation can be cast as a SN]

Description logics (CLASSIC, LOOM, OWL-DL, KAON, SROIQ, ...); [expressively inadequate]

Conceptual meaning representations (Schank's CD, Jackendoff, FrameNet, ...); [expressively/inferentially inadequate]

Thematic role representations (e.g., Palmer, Gildea, & Xue '10); [expressively/inferentially inadequate]

Abstract meaning representation (AMR) (Banarescu et al. '13); [inferentially inadequate]

Hobbs' “flat” representation (Hobbs '06); [conflates distinct types]

Structured English (e.g., MacCartney & Manning '09, Dagan et al. '08, ...); [ambiguous, inferentially inadequate]

Montague Grammar (English as logic) (e.g., Dowty '79, Chierchia & McConnellGinet '00); [unnecessarily complex, higher-order]

Extensional Montague fragments (e.g., McAllester & Divan '92, Artzi & Zettlemoyer '13); [expressively inadequate]

DCS trees (Liang et al. '11); [expressively inadequate]

Situation semantics (Reichenbach '47, Barwise & Perry '83); [abstruse, inferentially inadequate]

Episodic Logic (EL); [remedies many of these flaws; still brittle, lacks adequate KB]

may be no universal principles constraining the placement of quantifiers or modifiers—as long as there are systematic clues, be they syntactic or semantic, as to what phrases quantify or modify what other phrases; and propositional attitudes need not necessarily be expressed through explicit recursion, but might be expressed via sequenced sentences.

4 Episodic Logic and EPILOG

Episodic Logic (EL) (e.g., Hwang & Schubert 1994, Schubert & Hwang 2000, Schubert 2013) is a Montague-inspired, language-like, first-order, situational, intensional SR/KR with a small number of types. The types include those of FOL, but also nonstandard quantifiers with restrictors, predicate and sentence modifiers, predicate and sentence reification operators, quoted expressions, and operators allowing characterization of episodes (events, situations, etc.) by sentences. Thus EL allows for most of the semantic phenomena found in natural languages, including reference to all sorts of concrete and abstract individuals, and temporal or causal relations between complex events. Logical forms (LFs) are obtained compositionally from phrase structure trees, and are subsequently normalized (e.g., Schubert 2014); the following are two examples of the transduction from NL to EL.

John donated blood to the Red Cross.
 [John1 <past donate.v> (K blood.n) Red_Cross] {initial LF}
 (some e: [e before Now3])
 (some x: [x blood.n])
 [[John1 donate.v x Red_Cross1] ** e] {scoped LF}
 Skolemize, split:
 [E1.sk before Now3], [Blood1.sk blood.n],
 [[John1 donate.v Blood1.sk Red_Cross1] ** E1.sk]
Very few people still debate the fact that the earth is heating up
 {Final representation}: [Fact4.sk fact.n], [Fact4.sk =
 (that (some e0: [e0 at-about Now0]
 [(The z [z earth.n] [z heat_up.v]) ** e0])),
 ((fquan (very.adv few.a) x: [x (plur person.n)]
 (still.adv (l v [v debate.v Fact4.sk]))))

Currently this transduction is quite error-prone for sentences of any complexity, primarily because we start with Treebank parses that are often erroneous as well as underspecified from a semantic perspective (e.g., not distinguishing SBAR constituents that are relative clauses from ones that are adverbials or clausal nominals).

EPILOG is the inference engine for EL, and has existed in various forms and precursors for around 25 years (e.g., see Schubert & Hwang 2000). It consists of a core reasoner capable of goal-driven and input-driven inference, and is assisted by “general-purpose specialists” via a uniform interface. EPILOG 1 used specialists for taxonomies, partonomies, time, equality, arithmetic, and seven others, while in the current version, EPILOG 2, specialist integration remains incomplete.² EPILOG’s inference methods include a very general form of embedded resolution that doesn’t require clause form, as well as several natural deduction rules, such as assumption of the antecedent in proving a conditional.

It is often argued—speciously—that expressivity needs to be kept modest to assure efficient inference. But EPILOG 2 was shown to hold its own against highly optimized FOL engines on large FOL problems, despite its language-like richness (Morbini & Schubert ’09). The problems that remain are that inference is brittle, in the sense that there is no provision

²Development of complex AI systems by an academic PI collaborating with a few grad students tends to proceed in fits and starts!

for bridging knowledge gaps with assumption-making; uncertainty handling is heuristic rather than well-founded; and, of course, its KB remains inadequate for general understanding and commonsense reasoning: It is a rocket engine waiting for an adequate supply of fuel.

The following are some examples of EPILOG inference. First, here is an example of an inference based on the non-standard (but quite common) quantifier *most*. (The example is based loosely on James Allen’s and George Ferguson’s collection of human-human “Monroe domain” dialogues concerned with urban emergency response.)

Given: *Most front loaders are currently in use;*
 Background knowledge:
 Whatever equipment is in use is unavailable;
 Front loaders are equipment.
 Conclusion: Most front loaders are currently unavailable.
 Expressed in EL:
 (most x: [x front-loader] [[x in-use] @ Now3]);
 (all x: [x equipment])
 (all e [[x in-use] @ e] => [(not [x available]) @ e]]);
 (all x: [x front-loader] [x equipment])
 Conclusion by EPILOG:
 (most x: [x front-loader] [(not [x available]) @ Now3])

The next example illustrates reasoning involving attitudes (using English glosses of EL formulas for space reasons):

Given:
Alice found out that Mark Twain is the same as Samuel Clemens.
 Background knowledge (with “you” meaning anyone):
 When you find out something, you don’t know it to be true at the start (of the finding-out event) but know it to be true at the end.
 Whatever you know to be true is true.
 Conclusions:
 - Alice didn’t know (before finding out) that Mark Twain is the same as Samuel Clemens;
 - Alice knew afterwards that Mark Twain is the same as Samuel Clemens;
 - Mark Twain is the same as Samuel Clemens.

EPILOG’s method of embedded inference resembles Natural Logic (NLog) but is more general. An example beyond the scope of NLog, again from Allen and Ferguson’s Monroe domain (with the question shown explicitly but with the relevant facts left as English glosses—details can be found in (Schubert 2013)):

Given: *The small crane, which is on Clinton Ave, is not in use.*
 Background knowledge:
 Whatever equipment is not in use is available.
 Every available crane can be used to hoist rubble onto a truck.
 All cranes are equipment.
 Question: Can the small crane be used to hoist rubble from the collapsed building on Penfield Rd onto a truck?
 Formal query (as actually posed, without “prettified” syntax):
 (q (p (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))))))

EPILOG answered affirmatively in a fraction of a second.

5 Types of knowledge required

Assuming that EL, or some similar SR/KR, will be adequate for capturing knowledge that can be communicated in the form of sentences, as well as for combinatory reasoning that draws on a factual KB, let's consider some of the kinds of knowledge the examples we have seen call for.

Patterns of predication and modification: We saw how occurrence frequencies might account for disambiguation of *Time flies like an arrow*, and the same idea could be applied to PP attachment in sentences such as the one about seeing a bird with binoculars/yellow tail feathers. However, if we limit ourselves to the word level, the number of combinations that would have to be inventoried—and learned—would be astronomical. A more plausible approach is to learn *patterns of predication/modification*, abstracted from various occurrences. For example, the variants *John saw a warbler {with yellow feathers, with his opera glasses}* instantiate much the same patterns as the earlier examples, namely ones like *person see bird, see with viewing-instrument, bird with feathers* (or even *bird with animal-body-part*). From a ML point of view, these patterns would be features whose instantiation in text will favor the phrases that instantiate them.

Patterns of actions/events and relationships: For expanding and connecting sentences in a text into a semantically coherent discourse, for instance in the case of Schank's restaurant stories or the terminal cancer story, we need schemas or patterns that extend beyond mere patterns of predication and modification. However, in contrast with many historically proposed schemas, the propositional parts of the schemas should allow for the full expressivity of language. The following is a sketch of a schema for visiting a foreign country for pleasure (for clarity, some predicates have been written as English phrases):

Dyn-schema [Person x visits foreign-country y] (episode e):
 Init-conds: [x loc-at u:<some (place in (home-country-of y))>] e1
 Co-conds: [x have v:<some (travel paraphernalia for y)>] e2
 Steps: [x travel from u to w:<some (place in y)>] e3
 [x do some activities in y] e4
 [x travel from z:<some (place in y)> to u] e5
 Effects: [x obtain g:<some (gratification-from e4)>] e6
 Constraints: [e = (join-of e3 e4 e5 e6)], [e1 starts e],
 [e2 same-time e], [e3 consec e4 e5], [e6 during e4]

Similarly, we can imagine patterns of events experienced by cancer patients—diagnosis, surgery, chemotherapy, and so on. Also readers confronted with discourse about various object types think instantly of corresponding “object utility” patterns—what can you do with an apple? a pen? a car? The availability of corresponding schemas would lead immediately to tentative extrapolations for such sentences as *John picked up {a pen, an apple, a rental car}*. Similarly, schemas capturing “behavioral dispositions”—about circumstances where a dog will bark, a radar trap will lead to a citation, a fragile object will break, and so on. An important

characteristic of schemas (of the type advocated by Marvin Minsky, Roger Schank and others) is that schema-based inferences have a “match and extrapolate” (or abductive) character, rather than a logical one: If certain parts are matched, the corresponding instantiations of other parts become plausible hypotheses.

Lexical and paraphrase knowledge: Just to reiterate the flavor of this type of knowledge, here are some examples beyond those already mentioned: *Dogs* are land mammals; *pens* are writing instruments; *trees* are plants with a tall wooden trunk and a leafy branched crown; to *kill* is to render dead; to *walk* is to advance on foot; *managing* to do *x* entails doing *x*; *x sells y to z* \Leftrightarrow *z buys y from z*; *make good money* \Leftrightarrow *earn well* \Leftrightarrow *be well compensated* \Leftrightarrow *pull in the bucks*.

Conditional/generic world knowledge: Though much interesting work exists on generic and habitual sentences as “standalone” declaratives, they seem closely related to schemas, as characterized above. Consider for example the following generic claims: If you drop an object, it will hit whatever is directly beneath it; most dogs are someone's pet; dogs are generally friendly; chemotherapy tends to cause nausea and hair loss; etc. We surely possess tens of millions of such knowledge items; but they seem to evoke larger patterns of events and relationships, and as such are perhaps abstractions from, or summaries of, these larger schematic patterns.

Specialist knowledge: As mentioned earlier, EPILOG inference is supported by a number of specialists, and this seems essential for human-like facility in understanding and thinking about taxonomies, partonomies, temporal relations, arithmetic & scales, geometric/imagistic representations, sets, symbolic expressions (incl. language), and a few other pervasive sorts of entities and relations. Quantitatively, taxonomic and partonomic knowledge are undoubtedly very substantial, but the largest challenge may be the acquisition of geometric/imagistic representations of the objects, scenes and events in the everyday world.

6 Some attempts to address the “Knowledge Acquisition Bottleneck

Nowadays we have access to an abundance of textual “knowledge”, but regimenting this “knowledge” into a broad range of accurate pattern-like, schema-like, and axiom-like knowledge modules usable for understanding and reasoning remains a largely unsolved problem. Some believe that only an embodied learning approach can resolve the impasse, but I remain optimistic that some bootstrapping approach to learning from text, most likely beginning with several types of core knowledge, can reach the goal sooner.

I will elaborate slightly on the following past, present and planned efforts at the University of Rochester towards knowledge acquisition:

- KNEXT - knowledge extraction from text, aimed at general factoids (essentially, patterns of predication expressed as EL formulas) that could be used to guide a parser;
- “Sharpening” and abstracting of KNEXT factoids into quantified generalizations (Lore, successor to KNEXT);
- knowledge engineering (partially automated), especially for

- frequent and “primitive” verbs, several VerbNet classes, implicatives, and attitudinal verbs;
- WordNet gloss interpretation (in progress);
- graphics-based object/scene representation & inference;
- schema engineering, to assess syntactic/semantic requirements.

6.1 KNEXT – General knowledge extraction from text

General knowledge extraction using KNEXT begins with a Treebank-style parse. The parse tree is compositionally interpreted into EL using some 80 interpretive rules; these match regular expressions to the sequence of immediate constituents of a phrase, proceeding through the phrase structure tree in bottom-up, left-to-right fashion. At the same time a set of generalized factoids is derived; modifiers are used for factoid construction at the level at which they occur, but are dropped from the LFs at higher levels. Names are generalized to types such as *person*, *actor*, or *company*, and noun phrases are abstracted to the types corresponding to the head noun group. Various filters are applied to remove incoherent or otherwise undesirable factoids. For details see, e.g., (Van Durme & Schubert 2008), and references therein.

Roughly 200 million distinct factoids have been obtained in this way.³ Examples are (after automatic verbalization of the formulas) *A person may write a book*, *A computer may crash*, *A person may forget a password*, *Trees may have leaves*, etc. In evaluations, about 80% are judged to be reasonable general claims. Unfortunately, the 20% error rate is still too high for use of the factoids for effective parser guidance. We hope to curate the factoids via crowd-sourcing.

In their raw form, the factoids are also incapable of supporting inferences, apart from extremely weak ones such as that *Dana may have hair*, given the factoid that *A person may have hair* and that *Dana is a person*. However, about 6 million “sharpened” quantificational factoids have been derived from the original set, using tree transduction patterns, and semantic information from WordNet, VerbNet, and other sources (Gordon & Schubert 2010).⁴ The following shows an example of the effect of sharpening, and two inferences based on sharpened formulas.

E.g., A person may have hair →
 All or most persons permanently have some hair as part →
 (all-or-most x: [x person]
 (some e: [(x . e) permanent]
 (some y: [y hair]
 [[x have-as-part y] ** e]]))

Sample inferences:

E.g., Dana is a person →
 Probably, Dana permanently has some hair as part; i.e.,
 (probably
 (some e: [(Dana . e) permanent]
 (some y: [y hair]
 [[Dana has-as-part y] ** e]]))

E.g., ACME is a company (+ sharpened axiom not shown here) →

Probably, ACME occasionally announces a product; i.e.,
 (probably
 (occasional e
 (some y: [y product]
 [[ACME announce y] ** e]]))

These sorts of inferences seem capable of elaborating the properties of a given entity or situation; however, they fall short of providing a basis for understanding and reasoning.

6.2 Accumulating Lexical Knowledge: Primitives, VerbNet, WordNet, etc.

Our knowledge engineering efforts in this area were initially focused on axiomatizing around 150 “primitive” verbal concepts (such as (*move*, *grasp*, *see*, *learn*, *make*, *ask-of*, *convey-info-to*, *want-tbt*, ...), chosen for their occurrence frequency, utility in axiomatizing other verbs, and precedents in the literature. We also derived some 250 verb axioms from 15 VerbNet classes by providing a small number of axiom schemas for each class, where each schema contained predicate or modifier parameters that could be instantiated to provide verb-specific axioms. We paid particular attention to state-change verbs like *break*, *repair*, *melt*, etc., since these are particularly important in narrative understanding. Our axioms provide much more specific information than the semantic annotations in VerbNet itself. In a separate project we also created meta-axioms covering around 250 implicative and attitudinal verbs such as *manage (to)*, *neglect (to)*, *force (someone) (to)*, *learn (that)*, *know (that)*, etc. These can yield immediate factive and antifactive inferences that are crucial in discourse understanding.

An often-exploited source of informal lexical semantic information is the WordNet nominal hierarchy. We derived formalized axioms (some 77,000) for the majority of WN noun senses. Our work exploits the mass/count distinction and various other features of nouns (inferred from WordNet itself, VerbNet classes, and other sources) to correctly formalize the relation between hyponym-hypernym pairs that could easily be misanalyzed. For example, for the pair <seawater, water> the correct relation is that all seawater is water, whereas for the pair <gold, noble_metal> it would be incorrect to say that all gold is a noble metal; rather, it is the abstract kind, gold, that is an instance of a noble metal. Details of our lexical knowledge acquisition work can be found in (Stratos *et al.* 2011, Gordon & Schubert 2013, Schubert 2013) and references therein.

The coverage of our lexical axioms still remains too sparse to contribute decisively to any practical comprehension or reasoning tasks (which in any case depends as well on other kinds of knowledge, particularly schematic knowledge). We are currently working on fully interpreting verb-sense glosses in WordNet, which should take us another small step forward. A simple example is that from sense 2 of the verb *slam*, with gloss *to strike violently* and other associated information, we can derive an axiom to the effect that if [*x slam2.v y*] characterizes event *e*, then [*x violently strike1.v y*] also characterizes *e*, and *x* is probably a person and *y* is probably a thing (in sense 12, i.e., any physical entity). However, dictionary definitions are often only weakly informative, since they may be quasi-circular (e.g., *admire* is glossed as *feel admiration*

³See the browser at cs.rochester.edu/research/knext/browse/

⁴The Lore browser is at cs.rochester.edu/research/lore/browse/

for), may not apply to all synset members, and typically fail to provide major entailments (e.g., that *dying* results in being dead!)

Another recently initiated project is graphics-based scene modeling in first-reader stories. In one such story, we are told that Frank and Rosy see a nest in an apple tree. They wonder whether there are eggs in the nest and climb up the tree to find out. But to make sense of this, we need to understand why they wouldn't see the eggs while standing on the ground. The reason lies in visual occlusion of the eggs by the nest itself! This occlusion relation would be difficult to infer without a "visualization"; many such examples can be given, and this prompted our graphics-based approach (Bigelow *et al.* 2015).

The most recent project seeks to arrive at a generally adequate syntax for dynamic and relational schemas, along the lines indicated by the earlier sketch of a foreign holiday schema. In light of the earlier remarks about machine learning, we think this preparatory work is essential for eventual schema learning by reading, in a way that is less dependent on statistical co-occurrences in oceans of text, and more like the way children may learn about unfamiliar things and places from basic readers for their age group.

7 Concluding comments

Achieving broad, genuine language understanding and commonsense reasoning by machines will require both a clear conception of representational requirements, and major advances in capturing and encoding knowledge in those representations.

I have argued that we require a language-like propositional representation that is equally well-adapted to the semantic content of language and the content of general, commonsense world knowledge. The development of EL and the EPILOG inference engine indicate that such a representation can be designed and implemented. But though we have some understanding of reasoning in this framework (modulo well-founded uncertain inference), we are still grappling with the knowledge acquisition bottleneck.

Beyond a propositional representation, we also need "soft" knowledge consisting of millions of patterns of predication and modification that can be used to guide syntactic and semantic choices in the understanding process. And perhaps the most critical challenge is the acquisition of vast numbers of schemas (including imagistic ones) that delineate the stereotyped patterns of events and relationships in the world and that we recognize instantly in both linguistically and perceptually conveyed information.

Naturally, our efforts to acquire knowledge bases of high quality and adequate quantity will need to be coupled with efforts to gain a better grasp of the varieties of inference involved in human thinking – whether pattern-based, logical, schema-based, analogical, etc. This will in turn inform our understanding of knowledge representation requirements and perhaps show a way of breaking through the knowledge acquisition bottleneck.

References⁵

- E. Bigelow, D. Scarafoni, L. Schubert, and A. Wilson, "On the need for imagistic modeling in story understanding", *Biologically Inspired Cognitive Architectures 11*, Jan. 2015, 22–28. (See also BICA 2014, MIT.)
- G.N. Carlson and B. Spejowski, "Generic passages", *Natural Language Semantics 5*, 1997, 101–165.
- J. Gordon and L.K. Schubert, "WordNet hierarchy axiomatization and the mass-count distinction", *7th IEEE Int. Conf. on Semantic Computing (ICSC 2013)*, Irvine, CA, 2013.
- J. Gordon and L.K. Schubert, "Quantificational sharpening of commonsense knowledge", *Proc. of the AAAI Fall Symposium on Commonsense Knowledge*, Arlington, VA, Nov. 11-13, 2010.
- A.N. Kaplan and L.K. Schubert, "A computational model of belief", *Artificial Intelligence 120(1)*, June 2000, 119-160.
- F. Morbini and L.K. Schubert, "Evaluation of Epilog: A reasoner for Episodic Logic", *Commonsense'09*, June 1-3, 2009, Toronto, Canada.
- A. Purtee and L.K. Schubert, "TTT: A tree transduction language for syntactic and semantic processing", *EACL 2012 Workshop on Applications of Tree Automata Techniques in Natural Language Processing (ATANLP 2012)*, Avignon, France, Apr. 24, 2012.
- L.K. Schubert, "Semantic representation", *29th AAAI Conference (AAAI15)*, Jan. 25-30, 2015, Austin, TX.
- L.K. Schubert, "Computational linguistics", in Edward N. Zalta (Principal Editor), *Stanford Encyclopedia of Philosophy*, CSLI, Stanford, CA. (100pp), 2014.
- L.K. Schubert, "From Treebank parses to Episodic Logic and commonsense inference", *Proc. of the ACL 2014 Workshop on Semantic Parsing*, Baltimore, MD, June 26, ACL, 55-60.
- L.K. Schubert, "NLog-like inference and commonsense reasoning", in A. Zaenen, V. de Paiva, & C. Condoravdi (eds.), *Semantics for Textual Inference, special issue of Linguistic Issues in Language Technology (LiLT) 9(9)*, 2013.
- L.K. Schubert, "From generic sentences to scripts", *IJCAI'09 Workshop W22, Logic and the Simulation of Interaction and Reasoning (LSIR 2)*, Pasadena, CA, July 12, 2009.
- L.K. Schubert and C.H. Hwang, "Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding", in L. Iwanska and S.C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation*, MIT/AAAI Press, 2000, 111–174.
- K. Stratos, L.K. Schubert, & J. Gordon, "Episodic Logic: Natural Logic + reasoning", *Int. Conf. on Knowledge Engineering and Ontology Development, (KEOD'11)*, Paris, 2011.
- B. Van Durme and L.K. Schubert, "Open knowledge extraction using compositional language processing", *Symposium on Semantics in Systems for Text Processing (STEP 2008)*, September 22-24, 2008 - Venice, Italy.

⁵Apologies for the lack of broader referencing; extensive references can be found, e.g., in (Schubert 2013), listed above.