

From Treebank Parses to Episodic Logic and Commonsense Inference

Lenhart Schubert

Univ. of Rochester

`schubert@cs.rochester.edu`

1. Introduction and overview

We have developed an approach to broad-coverage semantic parsing that starts with Treebank parses and yields scoped, deindexed formulas in Episodic Logic (EL) that are directly usable for knowledge-based inference. Distinctive properties of our approach are

- the use of a tree transduction language, TTT, to partially disambiguate, refine (and sometimes repair) raw Treebank parses, and also to perform many deindexing and logical canonicalization tasks;
- the use of EL, a Montague-inspired logical framework for semantic representation and knowledge representation;
- allowance for nonclassical restricted quantifiers, several forms of modification and reification, quasi-quotes and syntactic closures;
- an event semantics that directly represents events with complex characterizations;
- a scoping algorithm that heuristically scopes quantifiers, logical connectives, and tense;
- a compositional approach to tense deindexing making use of tense trees; and
- the use of an inference engine, EPILOG, that supports input-driven and goal-driven inference in EL, in a style similar to (but more general than) Natural Logic.

We have applied this framework to general knowledge acquisition from text corpora and the web (though with tense meaning and many other semantic details stripped away) (e.g., Schubert & Hwang 2000, Van Durme & Schubert 2008), and more recently to caption interpretation for family photos, enabling alignment of names and other descriptors with human faces in the photos, and to interpreting sentences in simple first-reader stories. Ongoing projects are aimed at full interpretation

of lexical glosses and other sources of explicitly expressed general knowledge.

We now elaborate some of the themes in the preceding overview, concluding with comments on related work and important remaining challenges.

2. Refinement of Treebank parses using TTT

We generate initial logical forms by compositional interpretation of Treebank parses produced by the Charniak parser.¹ This mapping is encumbered by a number of difficulties. One is that current Treebank parsers produce many thousands of distinct expansions of phrasal categories, especially VPs, into sequences of constituents. We have overcome this difficulty through use of enhanced regular-expression patterns applied to sequences of constituent types, where our interpretive rules are associated directly with these patterns. About 100 patterns and corresponding semantic rules cover most of English.

Two other difficulties are that parsers still introduce about one phrasal error for every 10 words, and these can render interpretations nonsensical; and even when parses are deemed correct according to “gold standard” annotated corpora, they often conflate semantically disparate word and phrase types. For example, prepositional phrases (PPs) functioning as predicates are not distinguished from ones functioning as adverbial modifiers; the roles of *wh*-words that form questions, relative clauses, or *wh*-nominals are not distinguished; and constituents parsed as SBARs (subordinate clauses) can be relative clauses, adverbials, question clauses, or clausal nominals. Our approach to these problems makes use of a new tree transduction language, TTT (Purtee & Schubert 2012) that allows concise, modular, declarative representation of tree transductions. (As indicated below, TTT also plays a key role in logical form postprocessing.) While we cannot ex-

¹<ftp://ftp.cs.brown.edu/pub/nlparser/>

pect to correct the majority of parse errors in general texts, we have found it easy to use TTT for correction of certain systematic errors in particular domains. In addition, we use TTT to subclassify many function words and phrase types, and to partially disambiguate the role of PPs and SBARs, among other phrase types, allowing more reliable semantic interpretation.

3. EL as a semantic representation and knowledge representation

From a compositional perspective, the semantics of natural language is intensional and richly expressive, allowing for nonclassical quantifiers and several types of modification and reification. Yet many approaches to semantic interpretation rely on first-order logic (FOL) or some subset thereof as their target semantic representation. This is justifiable in certain restricted applications, grounded in extensional domains such as databases. However, FOL or description logics are often chosen as the semantic target even for broad-coverage semantic parsing, because of their well-understood semantics and proof theory and well-developed inference technology and, in some cases, by a putative expressiveness-tractability tradeoff. We reject such motivations – tools should be made to fit the phenomenon rather than the other way around. The tractability argument, for example, is simply mistaken: Efficient inference algorithms for subsets of an expressive representation can also be implemented within a more comprehensive inference framework, without forfeiting the advantages of expressiveness. Moreover, recent work in Natural Logic, which uses phrase-structured NL directly for inference, indicates that the richness of language is no obstacle to rapid inference of many obvious lexical entailments (e.g., MacCartney & Manning 2009).

Thus our target representation, EL, taking its cue from Montague allows directly for the kinds of quantification, intensionality, modification, and reification found in all natural languages (e.g., Schubert & Hwang 2000, Schubert, to appear). In addition, EL associates episodes (events, situations, processes) directly with arbitrarily complex sentences, rather than just with atomic predications, as in Davidsonian event semantics. For example, the initial sentence in each of the following pairs is interpreted as directly characterizing an episode, which then serves as antecedent for a

pronoun or definite:

*For many months, no rain fell;
this totally dried out the topsoil.*

Each superpower menaced the other with its nuclear arsenal; this situation persisted for decades.

Also, since NL allows for discussion of linguistic and other symbolic entities, so does EL, via quasi-quotation and substitutional quantification (closures). These can also express axiom schemas, and autocognitive reasoning (see further comments in Section 5).

4. Comprehensive scoping and tense deindexing

Though EL is Montague-inspired, one difference from a Montague-style intensional logic is that we treat noun phrase (NP) interpretations as unscoped elements, rather than second-order predicates. These elements are heuristically scoped to the sentence level in LF postprocessing, as proposed in (Schubert & Pelletier 1982). The latter proposal also covered scoping of logical connectives, which exhibit the same scope ambiguities as quantifiers. Our current heuristic scoping algorithm handles these phenomena as well as tense scope, allowing for such factors as syntactic ordering, island constraints, and differences in wide-scoping tendencies among different operators.

Episodes characterized by sentences remain implicit until application of a “deindexing” algorithm. This algorithm makes use of a contextual element called a *tense tree* which is built and traversed in accordance with simple recursive rules applied to indexical LFs. A tense tree contains branches corresponding to tense and aspect operators, and in the course of processing one or more sentences, sequences of episode tokens corresponding to clauses are deposited at the nodes by the deindexing rules, and adjacent tokens are used by these same rules to posit temporal or causal relations among “evoked” episodes. A comprehensive set of rules covering all tenses, aspects, and temporal adverbials was specified in (Hwang & Schubert 1994); the current semantic parsing machinery incorporates the tense and aspect rules but not yet the temporal adverbial rules.

Further processing steps, many implemented through TTT rules, further transform the LFs so as to Skolemize top-level existentials and definite NPs (in effect accommodating their presuppositions), separate top-level conjuncts, narrow

the scopes of certain negations, widen quantifier scopes out of episodic operator scopes where possible, resolve intrasentential coreference, perform lambda and equality reductions, and also generate some immediate inferences (e.g., inferring that *Mrs. Smith* refers to a married woman).

The following example, for the first sentence above, illustrates the kind of LF generated by our semantic parser (first in unscoped, indexical form, then the resulting set of scoped, deindexed, and canonicalized formulas). Note that EL uses predicate infixing at the sentence level, for readability; so for example we have (E0 BEFORE NOW0) rather than (BEFORE E0 NOW0). ‘**’ is the operator linking a sentential formula to the episode it characterizes (Schubert 2000). ADV-S is a type-shifting operator, L stands for λ , and PLUR is a predicate modifier that converts a predicate over individuals into a predicate over sets of individuals.

For many months, no rain fell;

Refined Treebank parse:

(S (PP-FOR (IN for) (NP (CD many) (NNS months))) (l,l,l)
(NP (DT no) (NN rain)) (VP (VBD fell)) (l:l,l:l))

Unscoped, indexical LF (keys :F, :P, etc., are dropped later):

(:F (:F ADV-S (:P FOR.P (:Q MANY.DET (:F PLUR MONTH.N)))
(:I (:Q NO.DET RAIN.N) (:O PAST FALL.V)))

Canonicalized LFs (without adverbial-modifier deindexing):

(MONTHS0.SK (PLUR MONTH.N)), (MONTHS0.SK MANY.A),
((ADV-S (L Y (Y FOR.P MONTHS0.SK)))
((NO Z (Z RAIN.N) (SOME E0 (E0 BEFORE NOW0)
(Z FALL.V))) ** E0)

With adverbial deindexing, the prefixed adverbial modifier would become a predication (E0 LASTS-FOR.V MONTHS0.SK); E0 is the episode of no rain falling and MONTHS0.SK is the Skolem name generated for the set of many months.

5. Inference using the EPILOG inference engine

Semantic parsers that employ FOL or a subset of FOL (such as a description logic) as the target representation often employ an initial “abstract” representation mirroring some of the expressive devices of natural languages, which is then mapped to the target representation enabling inference. An important feature of our approach is that (scoped, deindexed) LFs expressed in EL are directly usable for inference in conjunction with lexical and world knowledge by our EPILOG inference engine. This has the advantages of not sacrificing any of the expressiveness of language, of linking inference more directly to surface form (in prin-

ciple enabling incremental entailment inference), and of being easier to understand and edit than representations remote from language.

EPILOG’s two main inference rules, for input-driven (forward-chaining) and goal-driven (backward-chaining) inference, substitute consequences or anti-consequences for subformulas as a function of polarity, much as in Natural Logic. But substitutions can be based on world knowledge as well as lexical knowledge, and to assure first-order completeness the chaining rules are supplemented with natural deduction rules such as proof by contradiction and proof of conditional formulas by assumption of the antecedent.

Moreover, EPILOG can reason with the expressive devices of EL mentioned in Sections 1 and 3 that lie beyond FOL, including generalized quantifiers, and reified predicates and propositions. (Schubert, to appear) contains relevant examples, such as the inference from *Most of the heavy Monroe resources are located in Monroe-east*, and background knowledge, to the conclusion *Few heavy resources are located in Monroe-west*; and inference of an answer to the modally complex question *Can the small crane be used to hoist rubble from the collapsed building on Penfield Rd onto a truck?* Also, the ability to use axiom schemas that involve quasi-quotes and syntactic closures allows lexical inferences based on knowledge about syntactic classes of lexical items (i.e., meaning postulates), as well as various forms of metareasoning, including reasoning about the system’s own knowledge and perceptions (Morbini & Schubert 2011). Significantly, the expressiveness of EL/EPILOG does not prevent competitive performance on first-order commonsense knowledge bases (derived from Doug Lenat’s Cyc), especially as the number of KB formulas grows into the thousands (Morbini & Schubert 2009).

In the various inference tasks to which EPILOG was applied in the past, the LFs used for natural language sentences were based on presumed compositional rules, without the machinery to derive them automatically (e.g., Schubert & Hwang 2000, Morbini & Schubert 2011, Stratos et al. 2011). Starting in 2001, in developing our KNEXT system for knowledge extraction from text, we used broad-coverage compositional interpretation into EL for the first time, but since our goal was to obtain simple general “factoids”—such

as that *a person may believe a proposition, people may wish to get rid of a dictator, clothes can be washed, etc.* (expressed logically)—our interpretive rules ignored tense, many modifiers, and other subtleties (e.g., Van Durme & Schubert 2008).

Factoids like the ones mentioned are unconditional and as such not directly usable for inference, but many millions of the factoids have been automatically strengthened into quantified, inference-enabling commonsense axioms (Gordon & Schubert 2010), and allow EPILOG to draw conclusions from short sentences (Gordon 2014, chapter 6). An example is the inference from *Tremblay is a singer* to the conclusion *Quite possibly Tremblay occasionally performs (or performed) a song* (automatically verbalized from an EL formula). Here the modal and frequency modification would not easily be captured within an FOL framework.

Recently, we have begun to apply much more complete compositional semantic rules to sentences “in the wild”, choosing two settings where sentences tend to be short (to minimize the impact of parse errors on semantic interpretation): derivation and integration of caption-derived knowledge and image-derived knowledge in a family photo domain, and interpretation of sentences in first-reader stories. In the family photo domain, we have fully interpreted the captions in a small development set, and used an EPILOG knowledge base to derive implicit attributes of the individuals mentioned in the captions (by name or other designations). These attributes then served to align the caption-derived individuals with individuals detected in the images, and were subsequently merged with image-derived attributes (with allowance for uncertainty). For example, for the caption *Tanya and Grandma Lillian at her high school graduation party*, after correct interpretation of *her* as referring to Tanya, Tanya was inferred to be a teenager (from the knowledge that a high school graduation party is generally held for a recent high school graduate, and a recent high school graduate is likely to be a teenager); while Grandma Lillian was inferred to be a grandmother, hence probably a senior, hence quite possibly gray-haired, and this enabled correct alignment of the names with the persons detected in the image, determined via image processing to be a young dark-haired female and a senior gray-haired female respectively.

In the first-reader domain (where we are using

McGuffey (2005)), we found that we could obtain correct or nearly correct interpretations for most simple declaratives (and some of the stories consist entirely of such sentences). At the time of writing, we are still working on discourse phenomena, especially in stories involving dialogues. For example, our semantic parser correctly derived and canonicalized the logical content of the opening line of one of the stories under consideration,

Oh Rosie! Do you see that nest in the apple tree?

The interpretation includes separate speech acts for the initial interjection and the question. Our goal in this work is integration of symbolic inference with inferences from imagistic modeling (for which we are using the Blender open source software), where the latter provides spatial inferences such as that the contents of a nest in a tree are not likely to be visible to children on the ground (setting the stage for the continuation of the story).

Phenomena not handled well at this point include intersentential anaphora, questions with gaps, imperatives, interjections, and direct address (*Look, Lucy, ...*). We are making progress on these, by using TTT repair rules for phenomena where Treebank parsers tend to falter, and by adding LF-level and discourse-level interpretive rules for the resulting phrasal patterns. Ongoing projects are aimed at full interpretation of lexical glosses and other sources of explicitly expressed general knowledge. However, as we explain in the concluding section, we do not believe that full-fledged, deep story understanding will be possible until we have large amounts of general knowledge, including not only the kinds of “if-then” knowledge (about word meanings and the world) we and others have been deriving and are continuing to derive, but also large amounts of pattern-like, schematic knowledge encoding our expectations about typical object configurations and event sequences (especially ones directed towards agents’ goals) in the world and in dialogue.

6. Related work

Most current projects in semantic parsing either single out domains that assure highly restricted natural language usage, or greatly limit the semantic content that is extracted from text. For example, projects may be aimed at question-answering over relational databases, with themes such as geography, air travel planning, or robocup (e.g., Ge & Mooney 2009, Artzi & Zettlemoyer 2011,

Kwiatkowski et al. 2011, Liang et al. 2011, Poon 2013). Impressive thematic scope is achieved in (Berant et al. 2013, Kwiatkowski et al. 2013), but the target semantic language (for Freebase access) is still restricted to database operations such as join, intersection, and set cardinality. Another popular domain is command execution by robots (e.g., Tellex 2011, Howard et al. 2013, Artzi & Zettlemoyer 2013).

Examples of work aimed at broader linguistic coverage are Johan Bos' Boxer project (Bos 2008), Lewis & Steedman's (2013) CCG-Distributional system, James Allen et al.'s (2013) work on extracting an OWL-DL verb ontology from WordNet, and Draicchio et al.'s (2013) FRED system for mapping from NL to OWL ontology. Boxer² is highly developed, but interpretations are limited to FOL, so that the kinds of general quantification, reification and modification that pervade ordinary language cannot be adequately captured. The CCG-Distributional approach combines logical and distributional semantics in an interesting way, but apart from the FOL limitation, the induced cluster-based predicates lose distinctions such as that between *town* and *country* or between *elected to* and *ran for*. As such, the system is applicable to (soft) entailment verification, but probably not to reasoning. A major limitation of mapping natural language to OWL-DL is that the assertion component of the latter is essentially limited to atomic predications and their negations, so that ordinary statements such as *Most students who passed the AI exam also passed the theory exam*, or *If Kim and Sandy get divorced, then Kim will probably get custody of their children*, cannot be represented, let alone reasoned with.

7. Concluding thoughts

The history of research in natural language understanding shows two seemingly divergent trends: One is the attempt to faithfully capture the logical form of natural language sentences, and to study entailment relations based on such forms. The other is the effort to map language onto preexisting, schematic knowledge structures of some sort, intended as a basis for understanding and inference – these might be FrameNet-like or Minsky-like frames, concepts in a description logic, Schankian scripts, general plans as under-

stood in AI, Pustejovskyan telic event schemas, or something similar. Both perspectives seem to have compelling merits, and this leads us to suppose that deep understanding may indeed require both surface representations and schematic representations, where surface representations can be viewed as concise abstractions from, or summaries of, schema instances or (for generic statements) of the schemas themselves. But where we differ from most approaches is that we would want both levels of representation to support inference. The surface level should support at least Natural-Logic-like entailment inference, along with inference chaining – for which EL and EPILOG are well-suited. The schematic level would support “reasonable” (or default) expectations based on familiar patterns of events, actions, or relationships. Further, the schematic level should itself allow for language-like expressiveness in the specification of roles, steps, goals, or other components, which might again be abstractions from more basic schemas. In other words, we envisage hierarchically organized schemas whose constituents are expressed in a language like EL and allow for EPILOG-like inference. We see the acquisition of such schemas as the most pressing need in machine understanding. Without them, we are limited to either narrow or shallow understanding.

Acknowledgements

This work was supported by NSF grant IIS-0916599 and a subcontract to ONR STTR N00014-10-M-0297. The comments of the anonymous referees helped to improve the paper.

References

- J. Allen, J. Orfan, W. de Beaumont, C. M. Teng, L. Galescu, M. Swift, 2013. Automatically deriving event ontologies for a commonsense knowledge base. *10th Int. Conf. on Computational Semantics (IWCS 2013)*, Potsdam, Germany, March 19-22.
- Y. Artzi and L. Zettlemoyer, 2011. Bootstrapping Semantic Parsers from Conversations. *Empirical Methods in Natural Language Processing (EMNLP)*, Edinburgh, UK.
- Y. Artzi and L. Zettlemoyer, 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. In *Trans. of the Assoc. for Computational Linguistics (ACL)*.
- J. Berant, A. Chou, R. Frostig, P. Liang, 2013. Semantic parsing on Freebase from question-answer pairs.

²www.meaningfactory.com/bos/pubs/Bos2008STEP2.pdf

- Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, WA.
- J. Bos, 2008. Wide-coverage semantic analysis with Boxer. In J. Bos and R. Delmonte (eds.), *Semantics in Text Processing. STEP 2008 Conference Proceedings*. Research in Computational Semantics, College Publications, 277–286.
- F. Draicchio, A. Gangemi, V. Presutti, and A.G. Nuzozese, 2013. FRED: From natural language text to RDF and OWL in one click. In P. Cimiano et al. (eds.), *ESWC 2013, LNCS 7955*, Springer, 263–267.
- R. Ge and R. J. Mooney, 2009. Learning a Compositional Semantic Parser using an Existing Syntactic Parser, In *Proceedings of the ACL-IJCNLP 2009*, Suntec, Singapore.
- J. Gordon, 2014. Inferential Commonsense Knowledge from Text. Ph.D. Thesis, Department of Computer Science, University of Rochester, Rochester, NY. Available at <http://www.cs.rochester.edu/u/jgordon/>.
- J. Gordon and L. K. Schubert, 2010. Quantificational sharpening of commonsense knowledge. *Common Sense Knowledge Symposium (CSK-10)*, AAAI 2010 Fall Symposium Series, Arlington, VA.
- T. M. Howard, S. Tellex, and N. Roy, 2013. A Natural Language Planner Interface for Mobile Manipulators. *30th Int. Conf. on Machine Learning*, Atlanta, GA, JMLR: W&CP volume 28.
- C. H. Hwang and L. K. Schubert, 1994. Interpreting tense, aspect, and time adverbials: a compositional, unified approach. D. M. Gabbay and H. J. Ohlbach (eds.), *Proc. of the 1st Int. Conf. on Temporal Logic*, Bonn, Germany, Springer, 238–264.
- T. Kwiatkowski, E. Choi, Y. Artzi, and L. Zettlemoyer, 2013. Scaling semantic parsers with on-the-fly ontology matching. *Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, WA.
- T. Kwiatkowski, L. Zettlemoyer, S. Goldwater, and M. Steedman, 2011. Lexical generalization in CCG grammar induction for semantic parsing. *Empirical Methods in Natural Language Processing (EMNLP)*, Edinburgh, UK.
- M. Lewis and M. Steedman, 2013. Combined distributional and logical semantics. *Trans. of the Assoc. for Computational Linguistics 1*, 179–192.
- P. Liang, M. Jordan, and D. Klein, 2011. Learning dependency-based compositional semantics. *Conf. of the Assoc. for Computational Linguistics (ACL)*, Portland, OR.
- B. MacCartney and C.D. Manning, 2009. An extended model of natural logic. *8th Int. Conf. on Computational Semantics (IWCS-8)*, Tilburg University, Netherlands. Assoc. for Computational Linguistics (ACL), Stroudsburg, PA.
- W.H. McGuffey, 2005 (original edition 1879). *McGuffey's First Eclectic Reader*. John Wiley and Sons, New York.
- F. Morbini and L. K. Schubert, 2011. Metareasoning as an integral part of commonsense and autocognitive reasoning. In M.T. Cox and A. Raja (eds.), *Metareasoning: Thinking about Thinking*, Ch. 17. MIT Press, Cambridge, MA, 267–282.
- F. Morbini & Schubert, 2009. Evaluation of Epilog: A reasoner for Episodic Logic. Commonsense'09, June 1-3, Toronto, Canada. Available at <http://commonsensereasoning.org/2009/papers.html>.
- H. Poon, 2013. Grounded unsupervised semantic parsing. *51st Ann. Meet. of the Assoc. for Computational Linguistics (ACL)*, Sofia, Bulgaria.
- A. Purtee and L. K. Schubert, 2012. 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.
- L. K. Schubert, to appear. NLog-like inference and commonsense reasoning. In A. Zaenen, V. de Paiva and C. Condoravdi (eds.), *Semantics for Textual Inference*, CSLI Publications. Available at <http://www.cs.rochester.edu/u/schubert/papers/nlog-like-inference12.pdf>.
- L. K. Schubert, 2000. The situations we talk about. In J. Minker (ed.), *Logic-Based Artificial Intelligence*, Kluwer, Dordrecht, 407–439.
- L. K. Schubert and C. H. Hwang, 2000. 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: Language for Knowledge and Knowledge for Language*. MIT/AAAI Press, Menlo Park, CA, and Cambridge, MA, 111–174.
- L. K. Schubert and F.J. Pelletier, 1982. From English to logic: Context-free computation of 'conventional' logical translations. *Am. J. of Computational Linguistics 8*, 27–44. Reprinted in B.J. Grosz, K. Sparck Jones, and B.L. Webber (eds.), *Readings in Natural Language Processing*, Morgan Kaufmann, Los Altos, CA, 1986, 293–311.
- S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy, 2011. Understanding natural language commands for robotic navigation and mobile manipulation. *Nat. Conf. on Artificial Intelligence (AAAI 2011)*, San Francisco, CA.
- B. Van Durme and L. K. Schubert, 2008. Open knowledge extraction through compositional language processing. *Symposium on Semantics in Systems for Text Processing (STEP 2008)*, Venice, Italy.