

# Entailment Inference in a Natural Logic-like General Reasoner

**Lenhart K. Schubert**  
Department of Computer Science  
University of Rochester  
Rochester, New York 14627

**Benjamin Van Durme**  
HLTCOE  
Johns Hopkins University  
Baltimore, MD 21211

**Marzieh Bazrafshan**  
Department of Computer Science  
University of Rochester  
Rochester, New York 14627

## Abstract

Recent work on entailment suggests that natural logics are well-suited to determining whether one sentence lexically entails another. We show how the EPILOG reasoning engine, designed for a natural language-like meaning representation (Episodic Logic, or EL), can be used to emulate natural logic inferences, while also enabling more general inferences such as ones from multiple premises, or ones based on world knowledge. Thus, to exploit the capabilities of EPILOG, we are working to populate its knowledge base with the kinds of lexical knowledge on which natural logics rely.

## Introduction: Natural logic and EPILOG

An interesting recent development in the area of recognizing textual entailment (RTE) has been the application of *natural logics* (van Benthem 1991; Valencia 1991; van Eijck 2005; MacCartney and Manning 2008; 2009; MacCartney 2009). Natural logics use meaning representations that are essentially phrase-structured NL sentences, and compute entailments as sequences of substitutions for constituents (words or phrases). For example, the fact that the verb phrase “won the Nobel Peace Prize (NPP)” is a special case of “received an honor” is the basis of the following entailments:

Several heads of state have won the NPP  
Several heads of state have received an honor

Everyone who has received an honor mentions it now and then  
Everyone who has won the NPP mentions it now and then

where the direction of substitution depends on the polarity of the sentential environment: positive (upward entailing) in the first case, and negative (downward entailing) in the second.

MacCartney and Manning (2008) demonstrated that surprisingly accurate entailment judgements on FRACAS test instances (Cooper et al. 1996) can be obtained with their NATLOG system, and that some gains on Recognizing Textual Entailment Challenge (RTEC) suites are also possible. In statistical semantics as well, Bar-Haim and Dagan (2008) among others have moved towards structured sentences (with variablization of certain constituents) as a basis for (graded) entailment, and have pointed to the importance of exploiting entailments at the lexical level.

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

However, RTE is a limited task, still rather far removed from the sort of reasoning based on linguistically conveyed information that has long been an AI dream. Current RTE systems, including MacCartney & Manning’s, depend on being given a premise sentence and a hypothesis, and typically proceed by “aligning” the premise with the hypothesis (much as two sentences in different languages are aligned in machine translation), and then judging whether a sequence of “edits” (substitutions, insertions, deletions) leading from the premise to the hypothesis makes it likely that the premise entails the hypothesis. Van Benthem (2007) suggests that natural logic inferences may comprise the “easy” part of commonsense reasoning. But for practical general application, an inference system should be able to (A) reason *forward* from facts and *backward* from goals, rather than being restricted to entailment *recognition*; (B) exploit lexical knowledge about causal and event structure, such as that purchasing  $x$  leads immediately to ownership of  $x$ ; (C) reason from multiple premises, e.g., concluding that Obama is a US citizen, given that: he was born in Hawaii, Hawaii is part of the US, an event occurring in a part of a region occurs in that region, and being born in the US implies being a US citizen; and (D) use lexical knowledge in concert with world knowledge (also illustrated by the previous example).

We first show that the EPILOG reasoning system, which has been under sporadic development since 1990, is well-suited to exploiting the insights gained in the recent natural logic-based work on RTE, while being capable of much broader inference functionality. EPILOG is based on a natural logic-like representation, Episodic Logic (EL), with a Montague-inspired syntax that is very close to that of natural language, and with inference mechanisms that readily lend themselves to emulation of natural logic, as we will show. Moreover, EPILOG is designed for forward as well as goal-directed inference, and it can make use of general world knowledge along with lexical knowledge.

We then turn to our effort to populate a knowledge base suitable for use within EPILOG, with the lexical properties whose importance has been noted and exploited by MacCartney and Manning for RTE: monotonicity, factivity, implicativity, presupposition, subsumption, synonymy, opposition (antonymy), and exclusion (without exhaustion). We do not dwell here on acquiring monotonicity, factivity, and implicativity properties of various word classes such as quan-

tifiers, adverbs, and modal verbs. This is straightforward in comparison: the number of lexical items that create non-upward monotone (non-positive) environments for their arguments is relatively small, and the properties at issue can be uniformly asserted for these classes.<sup>1</sup> Also entailment and exclusion relations can be enumerated fairly easily for quantifiers, prepositions, and some other classes of words.

A greater challenge confronting lexical entailment inference lies in acquiring relations between pairs of open-class words (and stock phrases, idioms, etc.). WordNet (Fellbaum 1998) is an important, but rather unreliable and incomplete source. In a 2001 study of the accuracy of subsumption and exclusion relations obtained from WordNet, Kaplan and Schubert (2001) found that as many as 1 out of 3 of such relations could be erroneous from a logical perspective. Recent improvements, such as a distinction between class membership and class inclusion, more complete hypernym and antonym relations among adjectives, and disambiguated word senses in glosses (e.g., see Clark, Fellbaum, and Hobbs (2008)) have somewhat alleviated these shortcomings, but it remains desirable to base classification of relations between pairs of words on additional sources and features besides WordNet, in the manner of MacCartney (2009). His decision-tree classifier, trained on over 2,400 labeled examples, also used NomBank (Meyers et al. 2004) for derivational relations, an automatically generated thesaurus by Dekang Lin for distributional similarity, string features, light-word flags, etc.

The approach we will outline starts with distributional similarity clusters rather than a labeled training set, and applies a classifier to pairs of their members, with the goal of deriving a systematic base of axiomatic lexical knowledge.

## Suitability of EL for Entailment Inference

We describe here how MacCartney and Manning’s natural logic approach to entailment inferences can be recast within our EPILOG-based framework, thus demonstrating the promise of EPILOG as a relatively mature tool that could be applied within this exciting emerging subfield of NLP.

## Episodic Logic (EL) and EPILOG

All human languages have devices for expressing predication, boolean connectives, generalized quantifiers (e.g., “*Most children of school-age ...*”), modalities (such as necessity, intention, and belief), uncertainty, explicit and implicit event reference, predicate modification (“*very smart*”, “*nearly awake*”) and predicate and sentence reification (“*Salsa dancing is fun*”, “*That the mission failed is tragic*”).

These observations motivated the development of EL, starting two decades ago, as a natural intensional logic subsuming FOL and including all of the above expressive devices. The crucial advantage of such a representation as a logical form for interpreting language is that LF-computation (and conversely, English verbalization of logical forms) becomes much simpler than for more restricted

<sup>1</sup>In addition, some inventories exist, for instance in MacCartney’s NATLOG code.

representations, and no work-arounds are needed for capturing the content of quantified, modal, modified, and reified expressions, or for describing or relating complex events. For example, EPILOG can infer from “*Most dogs occasionally bark*” that if Rover is a dog, then Rover probably barks occasionally: this is not an inference available in FOL.

Moreover, no distinction needs to be made between semantic representation and knowledge representation – reasoning is itself done in EL (with “reasoning” understood broadly, allowing for both deduction and uncertain inference). These points are made in detail in the publications concerning EL: Schubert and Hwang (2000) provide an overview of EL, its implementation in the EPILOG inference engine, and further references; Hwang and Schubert (1994) provides a comprehensive approach to time adverbials, tense, and aspect; Schubert (2000) revises and thoroughly develops the formal semantics of EL’s event-characterization operator “\*\*\*”.

Apart from being able to solve problems in self-aware and commonsense reasoning (Morbin and Schubert 2007; 2008; 2009), the system has been shown to hold its own in scalable first-order reasoning against the best current FOL theorem provers, and has been applied in the past to some challenge problems in theorem proving (which was not a design goal), and to small but varied domains in story understanding, aircraft maintenance messages, transportation planning dialogs, and belief inference by simulative reasoning (again see Schubert and Hwang (2000) and the references therein). The architecture of EPILOG is portrayed in Figure 1.

Reasoning in the “EPILOG core” is both goal- and input-driven, and rests on embedded unification and substitution operations that take into account the polarity of the subexpressions being unified. The forward inference illustrated in the figure is one that is enabled by a general world knowledge item stating that if a car crashes into a tree, the driver may be hurt or killed.<sup>2</sup>

## Encoding natural logic inference in EL

MacCartney and Manning (2008) use the following as a running example for explaining natural logic-based entailment inference, noting that the example is “*admittedly contrived but ... compactly exhibits containment, exclusion, and implicativity*”:

Jimmy Dean refused to move without his jeans

James Dean didn’t dance without pants

EPILOG is able to solve this example in an elegant way, closely analogous to MacCartney and Manning’s approach. We merely need to state monotonicity, subsectiveness, subsumption and exclusion properties of the predicates and modifiers involved, and supply general, vocabulary-independent meta-axioms that support reasoning with these properties. In particular, the following is the premise of the above example as stated in EL:

[jimmy-dean refuse (ka ((adv-a (without (k jeans))) move))].

<sup>2</sup>Note that EL formulas are written in infix form, with the “subject” argument preceding the predicate. Note also the occurrence of the event characterization operator “\*\*\*”, connecting a sentence to the event (more generally, episode) characterized by the sentence.

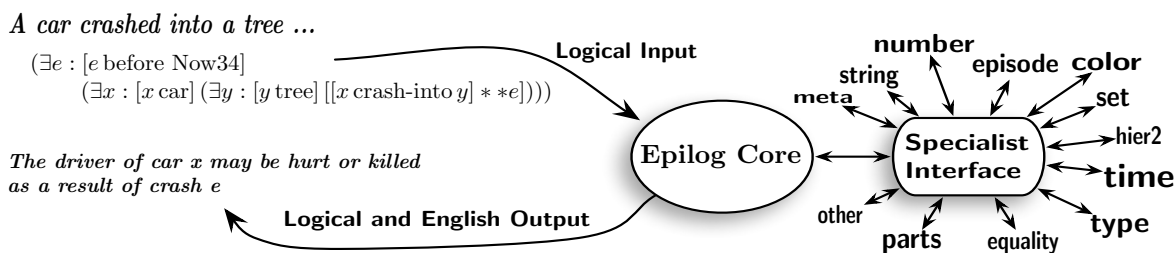


Figure 1: Episodic Logic and the EPILOG system

The background subsumption facts relating terms in the premise and conclusion are:

$(\text{all } x [[x \text{ jeans}] \Rightarrow [x \text{ pants}]]),$   
 $(\text{all } x [[x \text{ dance}] \Rightarrow [x \text{ move}]]),$

and the relevant lexical monotonicity property is:

[without monotone-pred2 1 -1].

This says that *without* is a 2-place predicate that is upward monotone in its subject argument and downward monotone in its object argument. Thus, in particular, being *without pants* entails *being without jeans*.

The implication of *refuse* assumed by MacCartney and Manning is that refusing to do something implicates not doing it, and this is formalized as in EL for use in EPILOG as:

$(\text{all\_pred } p (\text{all } x [[x \text{ refuse } (ka \text{ } p)] \Rightarrow (\text{not } [x \text{ } p])])).$

The quantifier 'all\_pred' indicates *substitutional* quantification over predicates, i.e., this is a formalized axiom *schema*, but is used directly for inference. The single quote is interpreted in EL as *quasi-quotation*, i.e., quotation that is opaque for object-level symbols but transparent for metavariables, in particular permitting substitution of object-level expressions for metavariables. The *ka* operator in the formula maps a monadic predicate to a *kind* of action or attribute.

The remaining facts are general meta-axioms: properties of equality, and half a dozen rules concerned with subsectivity of modifiers, the relationship of predicates to kinds, and very general monotonicity axioms such as:

$(\text{all\_pred } p [^p \text{ monotone-pred2 1 -1}]$   
 $(\text{all } x (\text{all } y (\text{all } z [[x \text{ } p \text{ } y] \text{ and } [z \text{ subkind } y] \text{ or } [z \text{ part-of } y]] \Rightarrow [x \text{ } p \text{ } z]])))).$

The axiom regarding *without* feeds into this general axiom, enabling (along with the remaining axioms) goal-directed inference of the conclusion in MacCartney and Manning's example. The answer is found in 4.5 seconds on a standard desktop computer running Allegro Common Lisp. An important point is that the conclusion is confirmed by a completely general reasoning mechanism that uses whatever axioms are in its knowledge base, rather than by an entailment *verifier*, dependent on being told from what premise it should derive the desired conclusion.

Just as in natural logic, EPILOG is not dependent on full sense disambiguation. For example, from "*Stock prices fell*" it follows that "*Stock prices went downward*", without providing any clear notion of what *stock prices* are, or what

sense of *fall* was intended. The only requirement would be an axiom that falling entails going downward.

### Inference in EL vs. (strict) natural logic

We do not claim that EL is a natural logic in the precise technical sense intended by its original proponents, notably Valencia (1991) and van Benthem (1991), and adhered to (and extended) by van Eijck (2005) and MacCartney and Manning (2008). The crucial shared properties are the close adherence to natural language syntax and semantics, and the use of replacement of embedded expressions by entailed, or equivalent expressions as inference mechanism.

The main difference is that monotonicity/ exclusion/ subsumption are not formalized cross-categorically in EL, but at the level of formulas. An advantage of the cross-categorical approach is that concisely storable relationships such as that 'dog' is subsumed by 'animal' or that 'extremely' is subsumed by 'very' can be used directly for transforming one formula into another, entailed formula. In EL, such facts are used to replace one embedded *sentence* by another. For example, instead of directly sanctioning replacement of predicate 'dog' by 'animal' in any upward entailing context, EL inference allows replacement of  $[x \text{ dog}]$  by  $[x \text{ animal}]$  (for any term  $x$ ) in any upward entailing sentential context. Similarly, instead of directly allowing replacement of 'extremely' by 'very' in any upward-entailing context, EL allows replacement of  $[x \text{ (extremely } P)]$  by  $[x \text{ (very } P)]$  in any upward-entailing sentential context.

Of course, we must specify that the predicates and other operators have these entailment properties. One way to do so is through axioms like:

$(\text{all } x [x \text{ dog}] [x \text{ animal}],$   
 $(\text{all\_pred } P (\text{all } x [x \text{ (extremely } P)] [x \text{ (very } P)])),$

but there are also more general ways to achieve the same effect, using EPILOG's type hierarchy and its meta-axiom capabilities. (We omit details for lack of space.) While this is a little more complex than the cross-categorical approach, an advantage of sentence-level entailment axiomatization is that it is more general, allowing for more than just relations between unitary elements such as 'dog' and 'animal', or 'extremely' and 'very' at the lexical level. For example, suppose that we want to establish that a certain individual, Brett, is a bachelor, given that Brett is a man and is not married. Axiomatically, we can assume the (approximate) definition:

$(\text{all } x [[x \text{ bachelor}] \Leftrightarrow [[x \text{ man}] \text{ and } (\text{not } [x \text{ married}])]]).$

But natural logic (in the forms so far proposed) does not

allow derivation of (Brett (is (a bachelor))) from:

(Brett (is (a man))), and (Brett (is (not married)));  
whereas [Brett bachelor] follows trivially in EL, given:  
[Brett man], and (not [Brett married]).

The strict adherence to surface form and avoidance of variables in (strict) natural logic is also a mixed blessing. Consider, for example, the role of reflexives as in the following entailment, not available in implemented versions of natural logic:

$\frac{(\text{Gosselin (thinks highly (of himself))})}{(\text{Someone (thinks highly (of Gosselin))})}$ ,

and while natural logics have used rules such as that modals block substitutions (e.g., *can run* does not entail *runs*), they currently do not account for positive inferences such as:

$\frac{(\text{Megan (learned (that ((Johnny Hart) (had (passed away))))))}{(\text{Megan (learned (that ((Johnny Hart) (had died))))})}$ ,

which is reasonable in this modal context because of the synonymy of *died* with *passed away* in the object of the mental attitude. These examples suggest that natural logics need extension to allow for various contextual parameters, not only constituent bracketing and polarity marking.<sup>3</sup>

In EL the first problem is avoided because the LF is computed as equivalent to “*Gosselin thinks highly of Gosselin*”, and the second can be dealt with using an axiom about substitution of lexically equivalent predicates in mental-attitude contexts. We have successfully run this example. At the same time EPILOG, informed that Johnny Hart was a famous cartoonist, would not infer that Megan learned that a famous cartoonist had died, as a person’s profession is not a matter of lexical knowledge, and thus might be unknown to Megan.

A further important advantage of the EL/EPILOG framework is (as the names suggest) its orientation towards events and situations in a changing world. For example, consider entailments (against a background of lexical knowledge about *acquire*) such as:

(past) event e : MS acquired Powerset

situation s right after e : MS owned Powerset

The temporal relationships involved (being in the past relative to the time of speech or writing; episodes immediately following one another) have generally been neglected in research on recognizing textual entailment, and they are not within the scope of current natural logics, yet are vitally important for understanding what a text is saying about the world. Moreover, most texts of practical interest are about events and situations in an ever-changing world, rather than about static relationships.

Finally, EL and EPILOG allow uniformly for inferences based on lexical knowledge and world knowledge (recall the earlier example of deducing Obama’s citizenship from his birth in Hawaii), which is not the case for systems like MacCartney and Manning’s NATLOG.

Commonsense reasoning performed by EPILOG using world knowledge was exemplified by Morbini and Schubert (2009), on questions such as “*Do pigs have wings?*”, “*Can you put out a fire with gasoline?*”, or “*Can you attack a person with a golf club?*”.<sup>4</sup>

<sup>3</sup>An observation also made by van Benthem (2007)

<sup>4</sup>The latter questions stem from the Cyc commonsense test suite: <http://www.cycfoundation.org/concepts/CommonSenseTest>.

## Forming a knowledge base of lexical relations

Based on the description provided by MacCartney (2009), we have reproduced the feature set of the natural logic relation classifier of MacCartney and Manning, for nominals, and are currently extending our implementation to cover additional parts of speech.<sup>5</sup> Rather than applying this classifier strictly online, as MacCartney and Manning did for RTE problems, we aim to develop a robust collection of axiomatized knowledge, using the following procedure:

1. Form word clusters based on distributional similarity.<sup>6</sup>
2. For each cluster, and each pair of words *v*, *w* in the cluster, apply a binary classifier to determine the relation between *v*, *w*, i.e., whether one entails the other, or they are synonyms, or one is the opposite of the other, or precludes the other (without exhausting all possibilities).
3. Prune “transitive edges” from entailment chains. E.g., if we get *cat*  $\models$  *mammal*, *mammal*  $\models$  *animal*, and *cat*  $\models$  *animal*, then we can discard the last relation, because it is a consequence of the first two.
4. Express the obtained information logically; e.g., the entailment between predicates *cat*, *mammal* becomes (all *x* [*x cat*] => [*x mammal*]), while (as was seen) the entailment between modifiers *extremely*, *very* becomes (all\_pred P (all *x* [[*x (extremely P)*] => [*x (very P)*]])); similarly the non-exhaustive exclusion for predicates *cat*, *dog* becomes (all *x* [*x cat*] (not [*x dog*])), while exclusion for modifiers *almost*, *entirely* becomes (all\_pred P (all *x* [[*x (almost P)*] => (not [*x (entirely P)*]]))).

Our preliminary results with this approach indicate that we can automatically classify the relationship between words in distributional similarity clusters with an accuracy ranging from 65% to over 90% (in many cases the figure is around 80%), depending on the word class and relationship at issue. Nominal entailment is very accurate because of the WordNet hypernym/hyponym paths exploited by some of the features in the classifier. Performance is worst for word pairs that are distant from each other in terms of WordNet paths, but our expectation is that these very indirect relations will eventually be discarded in favor of paths of direct relations (those retained in the pruning in step 3 above). A feature ablation study indicated that using either the Jiang-Conrath WordNet distance or the ‘DLin’ feature based on a thesaurus generated by Dekang Lin is critical, while other features such as string edit distance between lexemes and part-of-speech features for closed category words are less important (except for negation).

Since the logical formulas derived in step 4 will not be reliably true even after pruning, we expect to follow either

<sup>5</sup>A process significantly assisted by the Natural Language Toolkit (NLTK) of Bird, Klein, and Loper (2009).

<sup>6</sup>Clusters of words that tend to occur in the same contexts: same adjacent words, or words connected to the word of interest by a subject-of, argument-of, or other such features; such a resource has been acquired through the assistance of Patrick Pantel (p.c.).

of two strategies: further pruning, such as through Amazon Mechanical Turk; or attachment and use of certainty factors in the derived formulas. EPILOG 1 used probabilistic inference, and this will be re-implemented in the current EPILOG 2 system. Note also that probabilistic, inconsistency-tolerant methods such as Markov Logic Networks have become available and quite successful.

Beyond natural logic-like lexical axioms, we are exploring the use of existing, potentially error prone, collections of paraphrases rules for inference. For example, based on a small sample of positive cases from the second RTE challenge, we have estimated that roughly %50 of such items have their 'critical inference step' captured by a top-20 rewrite rule from the DIRT collection of Lin and Pantel (2001).

### Related work

First, though we have mentioned presuppositions, we have not characterized them or said how they affect entailment inference. As reiterated recently by Clausen and Manning (2009), a definitive characteristic of presuppositions lies in the way they project upward to the sentence level:

Bush [knew]/[didn't know] that Gore won the election.

If Gore won the election,

Bush [knew]/[didn't know] that Gore won the election.

Both the positive and negative version of the first sentence entail that Gore won the election, while this is not the case for either version of the second. For the first sentence, note the difference from an attitude with a non-sentential complement, such as "*Bush [remembers]/[doesn't remember] losing to Gore*" where the negation blocks the entailment that Bush lost to Gore. Thus Clausen and Manning project entailments and presuppositions separately, relying on a "plug/hole/filter" scheme (due to L. Karttunen) for the latter. In applying EPILOG, we would make this a matter of context-tracking, as most current theories of presupposition do. For example, in the second case above, the conditional form implicates that the speaker does *not* presume that Gore won the election, which prevents the presupposition from being added to the context. While "filters" deal with this case, they would not deal with the nonconditional case:

Maybe [Gore won the election]<sub>i</sub>, and maybe Bush knew it<sub>i</sub>,  
where the presupposition also is not accommodated.

The Nutcracker system for RTE (Bos and Markert 2006) has in common with our approach the use of formal logical representations. However, it relies on FOL theorem proving (with formulas derived from text and hypothesis DRS's and from WordNet relations), and as such is further removed from natural logic than what we are suggesting. In particular, no meta-axioms concerning modifiers, quantifiers, etc., can be used.

Yoad Winter and his collaborators have explored a number of technical issues in natural logic, such as the treatment of collective quantifiers (as in "*Two boys ate three pizzas*"), conjunction and relative clauses, and have proposed formal proof systems based on categorical grammar derivation trees, and the Lambek calculus (Winter 1998, Ben-Avia and

Winter 2004, Fyodorov et al. 2003, Zamansky et al. 2006). Their theoretical suggestions are interesting, and could be implemented as readily in EPILOG as in a natural logic.

A group at MITRE made an early attempt to employ the original EPILOG system to support entailment judgements in the first PASCAL RTE Challenge (Bayer et al. 2005). Results were decidedly suboptimal, for understandable reasons. "Flat" Davidsonian logical forms were employed that break a sentence into several smaller formulas, and beyond these LFs: "*very little additional semantic knowledge [was] exploited beyond a few added inference rules and simple word lists for semantic classification*". The rules included a rule for modals, such as that '*can run*' does not entail '*run*' and a rule for appositives. Applied to flat formulas, the rules produced errors (not due to EPILOG) such as that "*Rover is not a dog*" entails "*Rover is a dog*".

Beyond implementational issues, the primary obstacle was captured in Bayer et al.'s summary of EPILOG: *a rocket ship with nothing inside: fiendishly difficult to get off the ground, and unable to fly until a wide number of things work fairly well*. The full potential of EPILOG can only be realized with a large axiom set, formulated in the NL-like logical forms it is primarily designed for. It was difficult to see a few years ago how such axiom sets could be acquired, but as we have emphasized, the work on natural logic entailment has shed a great deal of light on what types of axioms are needed for effective entailment inference, and it is now clear that simple classificatory statements such as those above ([*'without monotone-pred2 1 -1*], [*'extremely stronger-mod-than 'very*], and the like), augmented with a small number of quantified meta-axioms (which we have also illustrated) will be sufficient to capture large classes of entailments of the sort addressed by MacCartney and Manning.

Another research current that is converging in some essential respects with natural logic is the statistically based work on textual entailment (Dagan *et al.* 2006). Though early approaches tended to focus on bag-of-words metrics and structural and lexical similarity, more recent work has pointed out the utility of viewing entailment inference as a proof system over parse trees (Bar-Haim and Dagan 2008). By allowing for variable tree nodes, the latter work is able to make systematic use of both manually coded rules (such as ones for conjunction) and learned ones such as:  $X \text{ bought } Y \rightarrow Y \text{ was sold to } X$ , which are reminiscent of the rules of Lin and Pantel. This is clearly a move towards NL-like logical form, similar to EL or natural logic formulas. Moreover, this line of work has recently begun to focus, like natural logic, on "sub-sentential entailment" based on lexical knowledge (as noted by Giampiccolo et al. (2007), and scrutinized in Mirkin et al. (2009)). The latter paper also points to the role of context polarity in the use of lexical entailment, further strengthening the link to natural logics. The authors note, for example, that we may replace *mammal* by *whale* in the restrictor of a universal quantifier, so that "*Mammals are warm-blooded*" entails "*Whales are warm-blooded*". They assert that "*this issue has rarely been addressed in applied inference systems*", though of course such inferences are elementary in logic-based systems, including EPILOG.

All this testifies to a convergence between originally very different perspectives on entailment, and underscores the need for an effective general inference engine that can easily be adapted to a variety of natural logic-like inferences; further, looking beyond entailment recognition, the engine should be capable of forward and backward inference, should cover causal and event semantics, and should allow for combination with world knowledge.

## Conclusion

We have shown through general considerations and examples that the EPILOG general inference system can emulate the types of inferences considered by recent work of MacCartney and Manning, in the context of RTE. The representation language, Episodic Logic, was described as being “natural logic-like”, while also being capable of supporting bidirectional reasoning, with use of multiple premises, attention to episodic and temporal relations, and integrated use of lexical and world knowledge. We briefly described ongoing efforts at incorporating lexical-semantic resources, in axiomitized form, for increasing the scope and robustness of EPILOG’s capabilities, as well as reproducing the binary classification mechanism of MacCartney and Manning.

Further work will enhance both EPILOG itself and its lexical and world knowledge base, with the goal of providing a general platform for knowledge-dependent, language-oriented inference, such as required for QA, helpful conversational and tutoring agents, and learning by reading.

**Acknowledgments** This work was supported in part by NSF awards IIS-0916599, and IIS-0910611.

## References

- Bar-Haim, R., and Dagan, I. 2008. Efficient semantic inference over language expressions. In *NSF Symposium on Semantic Knowledge Discovery, Organization and Use*.
- Bayer, S.; Burger, J.; Ferro, L.; Henderson, J.; and Yeh, E. 2005. Mitre’s Submissions to the EU Pascal RTE Challenge. In *First Challenge Workshop on Recognizing Textual Entailment*.
- Ben-Avi, G., and Winter, Y. 2004. Scope dominance with monotone quantifiers over finite domains. *Journal of Logic, Language and Information* 13:385–402.
- Bird, S.; Klein, E.; and Loper, E. 2009. *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*. O’Reilly Media.
- Bos, J., and Markert, K. 2006. When logical inference helps determining textual entailment (and when it doesn’t). In *Second PASCAL Challenges Workshop on Recognizing Textual Entailment*.
- Clark, P.; Fellbaum, C.; and Hobbs, J. 2008. Using and extending WordNet to support question-answering. In *4th Global WordNet Conference (GWC’08)*.
- Clausen, D., and Manning, C. D. 2009. Presupposed Content and Entailments in Natural Language Inference. In *ACL-IJCNLP Workshop on Applied Textual Inference*.
- Cooper, R.; Crouch, R.; van Eijck, J.; Fox, C.; van Genabith, J.; Jaspars, J.; Kamp, H.; Pinkal, M.; Milward, D.; Poesio, M.; Pulman, S.; Briscoe, T.; Maier, H.; and Konrad, K. 1996. Using the Framework. Technical report, FraCaS: A Framework for Computational Semantics.
- Dagan, I.; Glickman, O.; Gliozzo, A.; Marmorshtein, E.; and Strapparava, C. 2006. Direct Word Sense Matching for Lexical Substitution. In *ACL*.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*.
- Fyodorov, Y.; Winter, Y.; and Francez, N. 2003. Order-based inference in Natural Logic. *Logic Journal of the IGPL* 11:385–417.
- Giampiccolo, D.; Magnini, B.; Dagan, I.; and Dolan, B. 2007. The Third PASCAL Recognising Textual Entailment Challenge. In *ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*.
- Hwang, C., and Schubert, L. 1994. Interpreting tense, aspect, and time adverbials: a compositional, unified approach. In *1st International Conference on Temporal Logic*.
- Kaplan, A., and Schubert, L. 2001. Measuring and improving the quality of world knowledge extracted from WordNet. Technical Report 751, Dept. Computer Science, University of Rochester.
- Lin, D., and Pantel, P. 2001. DIRT - Discovery of Inference Rules from Text. In *SIGKDD*.
- MacCartney, B., and Manning, C. 2008. Modeling semantic containment and exclusion in natural language inference. In *COLING*.
- MacCartney, B., and Manning, C. 2009. An extended model of natural logic. In *IWCS-8*.
- MacCartney, B. 2009. *Natural Language Inference*. Ph.D. Dissertation, Stanford University.
- Meyers, A.; Reeves, R.; Macleod, C.; Szekely, R.; Zielinska, V.; Young, B.; and Grishman, R. 2004. The NomBank Project: An Interim Report. In *NAACL/HLT Workshop on Frontiers in Corpus Annotation*.
- Mirkin, S.; Dagan, I.; and Shnarch, E. 2009. Evaluating the inferential utility of lexical-semantic resources. In *EACL*.
- Morbini, F., and Schubert, L. 2007. Towards realistic autocognitive inference. In *AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*.
- Morbini, F., and Schubert, L. 2008. Metareasoning as an integral part of commonsense and autocognitive reasoning. In *AAAI Workshop on Metareasoning*.
- Morbini, F., and Schubert, L. 2009. Evaluation of Epilog: A reasoner for Episodic Logic. In *Commonsense*.
- Schubert, L., and Hwang, C. 2000. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In Iwańska, L., and Shapiro, S., eds., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*.
- Schubert, L. 2000. The situations we talk about. In Minker, J., ed., *Logic-Based Artificial Intelligence*. Kluwer, Dordrecht.
- Valencia, V. S. 1991. *Studies on Natural Logic and Categorical Grammar*. Ph.D. Dissertation, University of Amsterdam.
- van Benthem, J. 1991. Language in Action: categories, lambdas and dynamic logic. *Studies in Logic* 130.
- van Benthem, J. 2007. A brief history of natural logic. In *Int. Conf. on Logic, Navya-Nyaya & Applications: A Homage to Bimal Krishna Matilal*.
- van Eijck, J. 2005. Natural Logic for Natural Language. <http://homepages.cwi.nl/~jve/papers/05/nlnl/NLNL.pdf>.
- Winter, Y. 1998. *Flexible Boolean Semantics: coordination, plurality and scope in natural language*. Ph.D. Dissertation, Utrecht University.
- Zamansky, A.; Francez, N.; and Winter, Y. 2006. A ‘natural logic’ inference system using the lambek calculus. *Journal of Logic, Language and Information* 273–295.