EA NLU: Practical Language Understanding for Cognitive Modeling

Emmett Tomai and Kenneth D. Forbus

Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL, 60208, USA
{etomai, forbus}@northwestern.edu

Abstract

This paper presents an approach to creating flexible general-logic representations from language for use in high-level reasoning tasks in cognitive modeling. These representations are grounded in a large-scale ontology and emphasize the need for semantic breadth at the cost of syntactic breadth. The task-independent interpretation process allows task-specific pragmatics to guide the interpretation process. In the context of a particular cognitive model, we discuss our use of limited abduction for interpretation and show results of its performance.

Introduction

This paper describes our *practical language understanding* approach to facilitate natural language input to cognitive modeling experiments. These simulation experiments typically use materials that are adapted from prior experiments with human subjects, and many of the stimuli are natural language texts. They are very broad in terms of the topics that they use and the kinds of reasoning tasks that participants are asked to carry out. Thus they require large-scale knowledge and the ability to reason with knowledge that was originally provided in textual form. Typically the representations used as input for the simulations are created by hand from the original texts, a process that is both labor-intensive and error prone. It also leads to the problem of *tailorability*, since the simulation authors (or people working closely with them) do the encoding of the formal representations. By automating the process of converting natural language to formal representation, or even semi-automating it, tailorability is reduced, and the plausibility of the simulation results is increased.

We have implemented this approach in the Explanation Agent (EA) NLU system. EA NLU has been used in several cognitive modeling experiments including moral decision making (Dehghani et al, 2008), conceptual change (Friedman & Forbus, 2008) and blame attribution (Tomai & Forbus, 2008). Extra-linguistic cognitive modeling experiments provide a novel venue for natural language work. The models provide a precise and detailed definition of understanding in terms of inferential capability. The intersection of natural language phenomena and cognitively relevant understanding allows us to study computational language understanding in a rich context.

We start by introducing our practical language understanding approach to the challenges raised by cognitive simulations. We then present a specific cognitive model, MoralDM, to clarify the problem scope. We describe in detail the EA NLU semantic interpretation process, including the use of *limited evidential abduction*. Finally, we present experimental results on the performance of this abductive system showing evidence that it effectively controls complexity.

Practical Language Understanding

Our practical language understanding approach consists of three parts. First, a large knowledge base with an expressive representation language in necessary to support the depth and breadth of reasoning involved in cognitive modeling simulations. Second, a wide breadth of semantic forms must be supported from language, thus simplified syntax is used to make the challenge tractable. Third, the semantic interpretation process uses task-independent algorithms to provide a query-driven interface. This allows different models to assume different pragmatic contexts that drive the same semantic interpretation process.

For our knowledge base, we use the contents of ResearchCyc plus our own extensions, together around 2 million facts at present. This knowledge includes numerous *denotations* and *subcategorization frames* that link lexical terms to concepts in the Cyc ontology using *frame semantics* (Fillmore, 1982). These frames provide knowledge-rich semantics for words and common phrases. CycL is a very expressive predicate calculus representation language, including modals for handling belief and quotation, and microtheories to provide a logical environment for organizing and using knowledge. It supports higher-order expressions and makes no particular algorithmic commitments as to how they are handled.

For syntax, we use a simplified English grammar. Natural language understanding is about converting surface forms into some internal form and breadth can be considered along two dimensions. *Syntactic breadth* concerns the range of surface forms that can be processed. Most explorations of syntactic parsing have focused on maximizing syntactic breadth, but at the expense of impoverished internal forms. *Semantic breadth* concerns the range of ideas that can be expressed in the internal form, starting from natural language. For practical language understanding, we focus on semantic breadth instead of syntactic breadth. Having multiple ways to say the same thing makes an NL system easier to use (i.e. increase habitability, cf. Haas & Hendrix, 1980), but our goal is to maximize the number of things that can be said in terms of the underlying representations, not the surface forms. In earlier work (Kuehne & Forbus, 2004) we developed a simplified English grammar, QRGCE, that we extend as necessary to handle new tasks. Like other simplified languages, (e.g., Clark's CPL (Clark et al 2005)), it restricts grammar but does not *a priori* restrict the vocabulary, as controlled languages do. This enables most extensions to be made by adding vocabulary rather than changing the grammar.

For task-independent semantic interpretation, we use compositional frame semantics at the sentence-level and abductive back-chaining at the discourse-level. For each new sentence in a discourse, compositional semantics provide a fast, efficient way to build complex semantic representations from the Cyc subcategorization frames. Because it factors out context in the composition of each syntactic constituent, it is able to handle nested constructs without becoming computationally intractable. Ambiguities are generated but maintained in packed representations for later disambiguation. This context independent sentence-level representation is then interpreted in the context of the ongoing discourse with abductive back-chaining. Abductive reasoning has a higher complexity cost, but allows pragmatic context and world knowledge to guide and constrain disambiguation and reference resolution.

Moral Decision Making

EA NLU was used in recent work with MoralDM (Dehghani et al, 2008), a cognitive model which captures aspects of moral decision-making. An important phenomenon in moral decision making is the impact of so-called *sacred values* which can override utilitarian strategies. Consider the following scenario (from Ritov & Baron 1999):

> As a result of a dam on a river, 20 species of fish are threatened with extinction. By opening the dam for a month each year, you can save these species, but 2 species downstream will become extinct because of the changing water level.

Utilitarian reasoning argues for opening the dam, since it will save 20 species and only kill 2. But studies have found that US participants choose to not open it, even though more species would be saved. These results support the hypothesis that when a decision scenario involves values sacred to the decision-maker, the acceptability of actions becomes a more significant determining factor than the outcomes of those actions.

MoralDM accepts decision scenarios represented in predicate calculus and uses both first-principles and analogical reasoning to reach a decision. It was evaluated against psychological studies by Ritov and Baron (1999) and Waldmann and Dieterich (2007). Four scenarios from each study were tested. EA NLU was used to semi-automatically encode the scenarios into formal representation suitable for input to MoralDM. In all eight cases, MoralDM was able to make the decision that matched the human data. For more details on this experiment, see Dehghani et al. (2008).

Semantic Interpretation

EA NLU provides a query-driven, task-independent semantic interpretation process. A cognitive model (or other inferential task model) provides the pragmatic constraint necessary to guide this process by querying for task-specific facts. Here we will describe this process in the context of MoralDM. Consider the decision-making task for the dam scenario given above. It requires identifying entities, events and the role relations between them. It also requires resolving anaphoric references. The scenario describes causality and a set of hypothetical futures, one default and one contingent. The abstraction of choice between these futures is not explicitly mentioned but is key to the task of making a decision. The numeric quantification of the outcomes of action and inaction are important to evaluating a utilitarian decision, while world knowledge about environmental harm is necessary to appreciate the sacred aspect. These phenomena are expressed in the text of the scenario, and the EA NLU semantic interpretation process is able to construct appropriate representations for them. The pragmatic constraint of the cognitive model is necessary to disambiguate the numerous possible representations.

We now turn to the details of this process using the simplified English version of the dam scenario:

> Because of a dam on a river, 20 species of fish will be extinct. You can save them by opening the dam. The opening would cause 2 species of fish to be extinct.

The translation to simplified English is primarily conformation to the set of supported syntactic patterns. In addition, several details that do not directly impact MoralDM's understanding of the situation were omitted.

Compositional Frame Semantics

EA NLU uses Allen's bottom-up chart parser (Allen, 1994) with the COMLEX lexicon (Macleod et al. 1998) to produce a set of standard, hierarchical parse tree representations for a given sentence. Each constituent in the tree has a slot which holds a predicate calculus form representing the compositional semantics of that span of the input. At the leaf nodes of the tree, subcategorization frames from ResearchCyc are retrieved for word and common phrase semantics. These frames form the basis of grounding in the world knowledge of ResearchCyc. Because a given word is likely to have multiple possible interpretation frames, an explicit *choice set* is created from the set of frames. Figure 1 shows the (abbreviated) semantic form for the word "save" in the second sentence of our example scenario (the correct interpretation is the `RescuingSomeone` event). The capitalized colon-prefixed terms represent syntactic roles in the frames.

```
(choiceSet <identifier>
  (and (isa :ACTION SavingAFile)
       (informationOrigin :ACTION :OBJECT)
       (doneBy :ACTION :SUBJECT))
  (and (isa :ACTION RescuingSomeone)
       (beneficiary :ACTION :OBJECT)
       (performedBy :ACTION :SUBJECT))
  (and (isa :ACTION KeepingSomething)
       (performedBy :ACTION :SUBJECT)
       (objectActedOn :ACTION :OBJECT))
   …)
```

Figure 1: Predicate calculus semantics for the verb "save".

The parser uses a compositional lambda-calculus to merge constituent semantics and assign roles. Figure 2 shows the (abbreviated) semantic form for the constituent spanning the words "save them" in the second sentence of our example scenario. Based on the syntactic composition of a verb (save) and a pronoun (them), the :ACTION and :OBJECT roles have been assigned and the variables involved have been quantified.

```
(thereExists (TheList them29267 save29243)
 (choiceSet <identifier>
  (and (isa save29243 SavingAFile)
       (informationOrigin save29243 them29267)
       (doneBy save29243 :SUBJECT))
  (and (isa save29243 RescuingSomeone)
       (beneficiary save29243 them29267)
       (performedBy save29243 :SUBJECT))
  (and (isa save29243 KeepingSomething)
       (performedBy save29243 :SUBJECT)
       (objectActedOn save29243 them29267))
   …))
```

Figure 2: Predicate calculus semantics for the verb phrase "save them".

Following (Asher and Lascarides, 2003), we combine this compositional approach with a transformation process using dynamic logic principles from Discourse Representation Theory (DRT) (Kamp and Reyle, 1993). This process constructs a model-theoretic description of sentence content. Explicit quantifiers, negation and implication are handled according to DRT by constructing nested discourse representation structures (DRS). These DRS are represented in our logical environment using Cyc-style microtheories, with additional assertions for the universe of discourse variables. The universe reflects variable scoping while the particular embedding (implies, not, exactly, atLeast, etc) reflects the logical or numerical quantification of those variables. Figure 3 shows numerical quantification for the phrase "20 species of fish".

```
Universe: species29103

(isa species29103 Set-Mathematical)
(cardinality species29103 20)
(exactly 20 (DrsCaseFn DRS-3443637081-29216))

DRS-3443637081-29216:
  Universe: member-species29103 …

  (elementOf member-species29103 species29103)
  (isa member-species29103 BiologicalSpecies)
   …
```

Figure 3: Partial DRS for the phrase "20 species of fish".

The system uses this same representation for possible worlds indicated by modal statements with `possible` and `willBe` operators. Returning to our example, the second sentence involves a possible eventuality, saving the 20 species of fish by opening the dam. Figure 4 shows part of a DRS for the second sentence. Allowing for nested quantification and modal embedding in the model description gives the system considerable expressive power.

```
Universe: you29198

(possible (DrsCaseFn DRS-3443637928-29494))

DRS-3443637928-29494:

  Universe: open29299 save29243 them29267 …

  (isa save29243 RescuingSomeone)
  (performedBy save29243 you29198)
  (beneficiary save29243 them29267)
   …
```

Figure 4: Partial DRS for the sentence "You can save them by opening the dam."

Explicit Ambiguity

The DRS shown in figure 3 is only one of several possible representations of the second sentence in our example scenario. Ambiguities, such as the choice of frames shown in Figure 1, create multiple possibilities. Each ambiguity raised by the compositional semantics results in an explicit choice set. These reified sets are either open or closed with regard to additional choices. Parse tree choice sets are closed sets of the complete parse trees generated for a sentence. Frame semantics choice sets are closed sets of the available semantic frames retrieved from the knowledge base for a given term or span of terms in a sentence. Quantifier scope choice sets are closed sets of possible scoping configurations between two quantifiers or modal operators that are composed in the syntactic tree. Reference choice sets are open sets of possible referents for a referring term in a sentence. The set is open because the set of possible referents is not available to the sentence-level composition. At this time EA NLU supports pronominal and definite NP (including gerunds) anaphora.

Understanding by Abduction

Discourse understanding across multiple sentences is built by back-chaining from task-specific queries down to the sentence-level compositions. In this case the first-principles reasoning module in MoralDM issues queries about particular facts and abstractions salient to recognizing and making decisions. Figure 5 contains partial predicate calculus (variablized) for the abstraction of an action/inaction choice and its causal consequences.

```
(isa ?selecting SelectingSomething)

(choices ?selecting ?action)
(choices ?selecting ?inaction)

(isa ?inaction (InactionFn ?action))

(causes-PropSit
 (chosenItem ?selecting ?action) ?outcome))

(causes-PropSit
 (chosenItem ?selecting ?inaction) ?outcome2))
```

Figure 5: Predicate calculus query for a choice and its consequences.

Each of these queried facts is true if the facts in the sentence-level compositions provide the necessary antecedents to entail them given the domain theory axioms available. However, the truth of those antecedent facts and their (potentially nested) structure are dependent on the resolution of the choice sets. For any query, EA NLU must be able to identify whether there is a valid set of choices that entail the queried fact. To do this efficiently it uses abduction.

Several lines of research have explored understanding via abductive inference (Charniak & Goldman 1989, Ng & Mooney 1990, Hobbs 2004). Abductive proof is a very elegant and flexible framework, but it is under-constrained, and none of those efforts have tried to scale up to large knowledge bases. To make the problem tractable, we use limited evidential abduction. General abductive proof systems begin with the assumption that any fact in a proof may be assumed. To guide and control the search they use a cost-based heuristic such as path cost with weighted axioms, preference for consequents with partial antecedent support and graph interconnectedness within the proof. By contrast, our system begins by limiting abduction to only those cases where some manner of evidence outside the proof itself can be found to support the assumption. This is not mutually exclusive with internal heuristics, but at this point they have not been necessary. Instead, we use the explicit ambiguities created by the sentence compositions as a priori evidence. Each choice set is treated as a set of mutually exclusive reasonable assumptions. For example, consider the choice set for the word "save" shown in Figure 1. Given that ambiguity, in the nested DRS in Figure 4 (the possible eventuality) it may be freely assumed that there is one of: a `SavingAFile` event, a `RescuingSomeone` event or a `KeepingSomething` event together with their associated role relations (e.g. the `informationOrigin` fact for the `SavingAFile` event). In Figure 4, the truth of the `RescuingSomeone` fact is dependent on the choices that entail the `possible` nesting as well.

The compositional semantics of each sentence are expressed as a set of axioms that encode the dependencies between facts, choice sets and choices. These are combined with microtheories containing axioms and facts for the task and domain. When an abductive query is made for a set of facts such as those shown in Figure 5, an abductive proof is returned. This consists of bindings for the variables in the query and one or more sets of assumptions that entail those bindings.

Importantly, abduction uses the axiomization of the task to guide the search, eliminating the need to encode an additional set of heuristics for resolving particular ambiguities.

Evaluation

The evaluation of MoralDM described in (Dehghani et al. 2008) demonstrated the capability of EA NLU to meet the understanding requirements of the reasoning model. Each scenario from the source experiments was rendered in our simplified English and input to EA NLU. The system generated explicit ambiguities which were presented to the experimenter for manual disambiguation. Given this intervention, the system was able to produce logical representations sufficient for MoralDM to model human decision-making outcomes.

Here we present an evaluation of our limited evidential abduction for automatic disambiguation within these established constraints. The four scenarios from Ritov and Baron used in the prior experiment are used. The Waldmann and Dieterich scenarios were not rerun due to time constraints. Each scenario is taken in its simplified form and processed by EA NLU. The system then queries for the same set of facts that MoralDM queries for use in its first-principles reasoning module. The query is handled as an abductive proof which disambiguates the choice sets from the compositional semantics.

Table 1 shows the number of ambiguous choice sets (parse trees, frame semantics, quantifier scope and references) in each scenario. For the three closed sets, the average number of choices is given in parenthesis. Table 2 contains the same figures for unresolved choice sets after EA NLU builds the abductive proof. These are the choice sets that the system did not need to resolve in order to prove the necessary facts – they are considered spurious in the context of this task. Because there are constraints between choices, often the system reduced the available choices even when the set itself containing them was not resolved.

	Parse	Sem.	Scope	Ref.
Scenario1	3 (1.7)	13 (4.5)	5 (2)	4
Scenario2	2 (1)	15 (4)	5 (2)	5
Scenario3	3 (1)	6 (5.2)	3 (3)	4
Scenario4	3 (1)	13 (3.5)	3 (2)	5

Table 1: Explicit ambiguities (average number of choices per set given in parenthesis).

	Parse	Sem.	Scope	Ref.
Scenario1	0	5 (2.8)	1 (2)	1
Scenario2	0	6 (3.5)	2 (2)	2
Scenario3	0	1 (2)	0	1
Scenario4	0	2 (3.5)	0	2

Table 2: Unresolved ambiguities (average number of choices per set given in parenthesis).

In Table 3 we present the complexity space for each scenario. The worst-case number of random choices to satisfy the query is compared with the number of assumptions made by the abductive proof. The latter includes every time in the proof that the system checks to see if a fact can be or is already assumed. The space of unresolved choices is also provided. Figure 6 shows a graph of the total choices vs. the abductive assumptions.

	Total Choices	Abductive Assumptions	Unresolved Space
Scenario3	6.27×10^{4}	6.21×10^{3}	2
Scenario4	7.17×10^{6}	1.05×10^{5}	10
Scenario1	7.63×10^{7}	1.01×10^{4}	288
Scenario2	6.94×10^{8}	8.33×10^{4}	768

Table 3: Number of assumptions to prove the task-specific knowledge request.

In all four scenarios the abductive proof is able to provide the facts requested. The number of choice actions taken by the solver is between one to four orders of magnitude smaller than the space of possible choices. What is most notable though is that as the space increases, the number of unresolved (task-irrelevant) choices increases while the number of assumptions does not. This demonstrates that this approach is able to make the necessary choices without suffering as the number of unnecessary choices increases. In discourse processing this is particularly important as each additional sentence increases complexity regardless of whether the added ambiguities are task-relevant.

The types of ambiguities that the system did not resolve were largely surface distinctions in entity types. There were, for example, several ways of representing "species of fish" that did not impact this decision making task. Almost all scoping ambiguities were resolved by the system. Since hypothetical futures were a central part of understanding the decision, this is not unexpected.

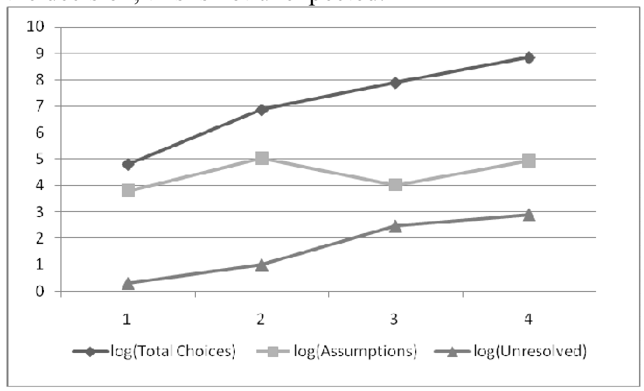

Figure 6: Complexity trends for abduction.

Related Work

Early work in deep semantic understanding (cf. Wilensky, 1981) demonstrated the necessity and power of rich knowledge in understanding, particularly the use of expectations. However, those early systems were brittle and not easily scalable, due in part to researchers having to generate all of the representations themselves. Subsequent work on robust, statistical methods focused on scalability and breadth, at the expense of depth. Common evaluation metrics (cf. ACE, TREC, MUC) capture only a few aspects of understanding. Recent web-scale knowledge extraction efforts (cf. Zamir & Etzioni, 1998) demonstrate scalability, but at the cost of limiting consideration to a small set of patterns for types of facts. These facts are clusters of word triples rather than formally expressed knowledge suitable for inference.

Recent investigations of deep understanding have either focused on formal arguments involving hand-generated examples, or focused on a single specific task. For example, most spoken-language dialogue systems (cf. Allen et al. 2007) are tuned for a specific type of task. Project Moebius (Friedland et al, 2006) focuses on knowledge capture in AP science domains using textbook-style assertions regarding facts and implications.

The Boxer (Bos et al, 2004) system, with the C&C Tools parser (Curran et al 2007), also creates DRT-style representations with concern for quantification and semantic role filling. However, it does not support future/hypothetical modalities nor does it ground its predicates in an ontology (or otherwise axiomize them) for general reasoning. We see this as a complementary effort, aiming to provide robust, large-scale processing working from the bottom up.

Conclusion

We have presented our practical language understanding approach to rich semantic understanding of natural language for cognitive modeling experiments. We use semantics grounded in a large-scale knowledge base and support complex quantification and modal operators. The semantic interpretation process combines the efficiency of compositional, context independent processing with pragmatically driven abductive back-chaining. We have discussed this in the context of MoralDM, a cognitive model of moral decision-making. In a key part of the semantic interpretation process, our EA NLU system explicitly creates choice sets which define the space of disambiguation for a given sentence. We have presented the results of experiments on the empirical performance of limited evidential abduction in disambiguating these choice sets for automatic encoding of moral decision making scenarios. These results give evidence that evidential abduction is an effective framework for automated interpretation.

Future Work

We intend to move forward with evidential abduction by integrating sources of evidence beyond EA NLU's explicit choice sets. We are looking at additional ways to leverage world knowledge in Cyc to provide evidence for assumptions based on domain-specific axioms. We are also exploring the use of narrative pragmatics as a constraint for story understanding. Part of that work will address how pragmatic expectations contribute evidence to abductive reasoning over linguistic ambiguity.

Ongoing projects are using EA NLU with cognitive models of conceptual change, story understanding, tutoring dialogue about commonsense knowledge and multimodal learning from science textbooks.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research.

References

Allen, J. F. 1994. *Natural Language Understanding*. (2nd ed). Redwood City, CA.: Benjamin/Cummings

Allen, J., Chambers, N., Ferguson, G., Galescu, L., Jung, H., Swift, M., and Taysom, W. 2007. PLOW: A Collaborative Task Learning Agent. *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07)*. Vancouver, Canada, Jul 22-26.

Asher, N. and Lascarides, A.. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge, UK.

Bos, J, Clark, S, Steedman, M, Curran, J and Hockenmaier, J (2004): Wide-Coverage Semantic Representations from a CCG Parser. *Proceedings of the 20th International Conference on Computational Linguistics* (COLING-04).

Charniak, E. and Goldman, R.P. 1989. A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. *IJCAI*.

Clark, P., Harrison, P., Jenkins, T., Thompson, J., and Wojcik, R. 2005. Acquiring and Using World Knowledge using a Restricted Subset of English. In *Proceedings of The 18th International FLAIRS Conference*.

Curran, J.R., Clark, S., and Bos, J. (2007): Linguistically Motivated Large-Scale NLP with C&C and Boxer. *Proceedings of the ACL 2007 Demonstrations Session* (ACL-07 demo).

Dehghani, M., Tomai, E., Forbus, K., Klenk, M. (2008). An Integrated Reasoning Approach to Moral Decision-Making. To appear in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI)*. Chicago, IL.

Friedland, N., Israel, D., Clark, P., Hovy, E., Hobbs, J., Mulkar, R., Porter, B., Barker, K., Weischedel, R., Martin, P. 2006. Project Möbius: A study on the feasibility of Learning by Reading. DISTAR Case # 11669.

Friedman, S. & Forbus, K. (2008). Learning Causal Models via Progressive Alignment & Qualitative Modeling: A Simulation. To appear in *Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci)*. Washington, D.C.

Haas, N. and Hendrix, G.G. 1980. An Approach to Acquiring and Applying Knowledge. Technical note ADA458593. SRI International Menlo Park, CA Artificial Intelligence Center.

Hobbs, J.R. 2004. Abduction in Natural Language Understanding. In Horn, L. and Ward, G. (eds.), *Handbook of Pragmatics*, pp. 724-741, Blackwell.

Kamp, H. and Reyle, U. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language*. Kluwer Academic Dordrecht; Boston.

Kuehne, S. E. and Forbus, K. (2004). On the Representation of Physical Quantities in Natural Language Text. *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society*, Chicago, Illinois, USA, August.

Macleod, C., Grishman, R., and Meyers, A. 1998. *COMLEX Syntax Reference Manual, Version 3.0*. Linguistic Data Consortium. University of Pennsylvania: Philadelphia, PA.

Ng, H.T. and Mooney, R.J. 1990. On the Role of Coherence in Abductive Explanation. *Proceedings of the Eighth National Conference on Artificial Intelligence*.

Ritov, I. and Baron, J. 1999. Protected Values and Omission Bias. *Organizational Behavior and Human Decision Processes*, 79(2): 79-94.

Waldmann, M. R., and Dieterich, J. 2007. Throwing a Bomb on a Person Versus Throwing a Person on a Bomb: Intervention Myopia in Moral Intuitions. *Psychological Science*, 18 (3), 247-253.

Wilensky, R. 1981. PAM and Micro PAM. In Roger C. Schank, & Christopher K. Riesbeck (Eds.), *Inside computer understanding*. Hillsdale, NJ: Erlbaum.

Zamir, O. and Etzioni, O. (1998). Web document clustering: a feasibility demonstration. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*.