# Recognizing Expressions of Commonsense Psychology in English Text

**Andrew Gordon, Abe Kazemzadeh, Anish Nair and Milena Petrova**
University of Southern California
Los Angeles, CA 90089 USA
gordon@ict.usc.edu, {kazemzad, anair, petrova}@usc.edu

## Abstract

Many applications of natural language processing technologies involve analyzing texts that concern the psychological states and processes of people, including their beliefs, goals, predictions, explanations, and plans. In this paper, we describe our efforts to create a robust, large-scale lexical-semantic resource for the recognition and classification of expressions of commonsense psychology in English Text. We achieve high levels of precision and recall by hand-authoring sets of local grammars for commonsense psychology concepts, and show that this approach can achieve classification performance greater than that obtained by using machine learning techniques. We demonstrate the utility of this resource for large-scale corpus analysis by identifying references to adversarial and competitive goals in political speeches throughout U.S. history.

## 1 Commonsense Psychology in Language

Across all text genres it is common to find words and phrases that refer to the mental states of people (their beliefs, goals, plans, emotions, etc.) and their mental processes (remembering, imagining, prioritizing, problem solving). These mental states and processes are among the broad range of concepts that people reason about every day as part of their commonsense understanding of human psychology. Commonsense psychology has been studied in many fields, sometimes using the terms *Folk psychology* or *Theory of Mind*, as both a set of beliefs that people have about the mind and as a set of everyday reasoning abilities.

Within the field of computational linguistics, the study of commonsense psychology has not received special attention, and is generally viewed as just one of the many conceptual areas that must be addressed in building large-scale lexical-semantic resources for language processing. Although there have been a number of projects that have included concepts of commonsense psychology as part of a larger lexical-semantic resource, e.g. the Berkeley FrameNet Project (Baker et al., 1998), none have attempted to achieve a high degree of breadth or depth over the sorts of expressions that people use to refer to mental states and processes.

The lack of a large-scale resource for the analysis of language for commonsense psychological concepts is seen as a barrier to the development of a range of potential computer applications that involve text analysis, including the following:

Natural language interfaces to mixed-initiative planning systems (Ferguson & Allen, 1993; Traum, 1993) require the ability to map expressions of users beliefs, goals, and plans (among other commonsense psychology concepts) onto formalizations that can be manipulated by automated planning algorithms.

Automated question answering systems (Voorhees & Buckland, 2002) require the ability to tag and index text corpora with the relevant commonsense psychology concepts in order to handle questions concerning the beliefs, expectations, and intentions of people.

Research efforts within the field of psychology that employ automated corpus analysis techniques to investigate developmental and mental illness impacts on language production, e.g. Reboul & Sabatier s(2001) study of the discourse of schizophrenic patients, require the ability to identify all references to certain psychological concepts in order to draw statistical comparisons.

In order to enable future applications, we undertook a new effort to meet this need for a linguistic resource. This paper describes our efforts in building a large-scale lexical-semantic resource for automated processing of natural language text about mental states and processes. Our aim was to build a system that would analyze natural language text and recognize, with high precision and recall, every expression therein related to commonsense psychology, even in the face of an extremely broad range of surface forms. Each recognized expression would be tagged with an appropriate concept from a broad set of those that participate in our commonsense psychological theories.

Section 2 demonstrates the utility of a lexical-semantic resource of commonsense psychology in automated corpus analysis through a study of the changes in mental state expressions over the course of over 200 years of U.S. Presidential State-of-the-Union Addresses. Section 3 of this paper describes the methodology that we followed to create this resource, which involved the hand authoring of local grammars on a large scale. Section 4 describes a set of evaluations to determine the performance levels that these local grammars could achieve and to compare these levels to those of machine learning approaches. Section 5 concludes this paper with a discussion of the relative merits of this approach to the creation of lexical-semantic resources as compared to other approaches.

## 2 Applications to corpus analysis

One of the primary applications of a lexical-semantic resource for commonsense psychology is toward the automated analysis of large text corpora. The research value of identifying commonsense psychology expressions has been demonstrated in work on children slanguage use, where researchers have manually annotated large text corpora consisting of parent/child discourse transcripts (Barsch & Wellman, 1995) and children sstorybooks (Dyer et al., 2000). While these previous studies have yielded interesting results, they required enormous amounts of human effort to manually annotate texts. In this section we aim to show how a lexical-semantic resource for commonsense psychology can be used to automate this annotation task, with an example not from the domain of children slanguage acquisition, but rather political discourse.

We conducted a study to determine how political speeches have been tailored over the course of U.S. history throughout changing climates of military action. Specifically, we wondered if politicians were more likely to talk about goals having to do with conflict, competition, and aggression during wartime than in peacetime. In order to automatically recognize references to goals of this sort in text, we used a set of local grammars authored using the methodology described in Section 3 of this paper. The corpus we selected to apply these concept recognizers was the U.S. State of the Union Addresses from 1790 to 2003. The reasons for choosing this particular text corpus were its uniform distribution over time and its easy availability in electronic form from Project Gutenberg (www.gutenberg. net). Our set of local grammars identified 4290 references to these goals in this text corpus, the vast majority of them begin references to goals of an adversarial nature (rather than competitive). Examples of the references that were identified include the following:

> They sought to use the rights and privileges they had obtained in the United Nations, to frustrate its purposes [*adversarial-goal*] and cut down its powers as an effective agent of world progress. (*Truman, 1953*)
>
> The nearer we come to vanquishing [*adversarial-goal*] our enemies the more we inevitably become conscious of differences among the victors. (*Roosevelt, 1945*)
>
> Men have vied [*competitive-goal*] with each other to do their part and do it well. (*Wilson, 1918*)
>
> I will submit to Congress comprehensive legislation to strengthen our hand in combating [*adversarial-goal*] terrorists. (*Clinton, 1995*)

Figure 1 summarizes the results of applying our local grammars for adversarial and competitive goals to the U.S. State of the Union Addresses. For each year, the value that is plotted represents the number of references to these concepts that were identified per 100 words in the address. The interesting result of this analysis is that references to adversarial and competitive goals in this corpus increase in frequency in a pattern that directly corresponds to the major military conflicts that the U.S. has participated in throughout its history.

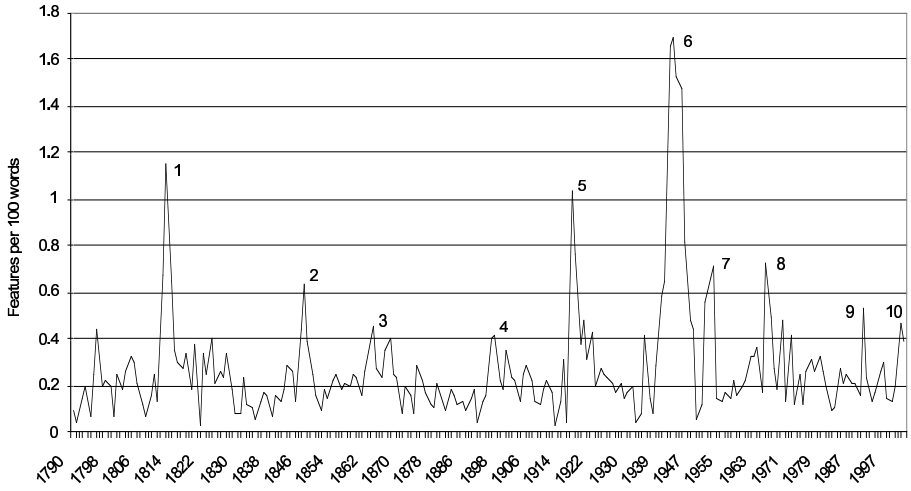


Figure 1. Adversarial and competitive goals in the U.S. State of the Union Addresses from 1790-2003

Each numbered peak in Figure 1 corresponds to a period in which the U.S. was involved in a military conflict. These are: 1) 1813, War of 1812, US and Britain; 2) 1847, Mexican American War; 3) 1864, Civil War; 4) 1898, Spanish American War; 5) 1917, World War I; 6) 1943, World War II; 7) 1952, Korean War; 8) 1966, Vietnam War; 9) 1991, Gulf War; 10) 2002, War on Terrorism.

The wide applicability of a lexical-semantic resource for commonsense psychology will require that the identified concepts are well defined and are of broad enough scope to be relevant to a wide range of tasks. Additionally, such a resource must achieve high levels of accuracy in identifying these concepts in natural language text. The remainder of this paper describes our efforts in authoring and evaluating such a resource.

## 3 Authoring recognition rules

The first challenge in building any lexical-semantic resource is to identify the concepts that are to be recognized in text and used as tags for indexing or markup. For expressions of commonsense psychology, these concepts must describe the broad scope of people smental states and processes. An ontology of commonsense psychology with a high degree of both breadth and depth is described by Gordon (2002). In this work, 635 commonsense psychology concepts were identified through an analysis of the representational requirements of a corpus of 372 planning strategies collected from 10 real-world planning domains. These concepts were grouped into 30 conceptual areas, corresponding to various reasoning functions, and full formal models of each of these conceptual areas are being authored to support automated inference about commonsense psychology (Gordon & Hobbs, 2003). We adopted this conceptual framework in our current project because of the broad scope of the concepts in this ontology and its potential for future integration into computational reasoning systems.

The full list of the 30 concept areas identified is as follows: 1) Managing knowledge, 2) Similarity comparison, 3) Memory retrieval, 4) Emotions, 5) Explanations, 6) World envisionment, 7) Execution envisionment, 8) Causes of failure, 9) Managing expectations, 10) Other agent reasoning, 11) Threat detection, 12) Goals, 13) Goal themes, 14) Goal management, 15) Plans, 16) Plan elements, 17) Planning modalities, 18) Planning goals, 19) Plan construction, 20) Plan adaptation, 21) Design, 22) Decisions, 23) Scheduling, 24) Monitoring, 25) Execution modalities, 26) Execution control, 27) Repetitive execution, 28) Plan following, 29) Observation of execution, and 30) Body interaction.

Our aim for this lexical-semantic resource was to develop a system that could automatically identify every expression of commonsense psychology in English text, and assign to them a tag corresponding to one of the 635 concepts in this ontology. For example, the following passage (from William Makepeace Thackeray s 1848 novel, *Vanity Fair*) illustrates the format of the output of this system, where references to commonsense psychology concepts are underlined and followed by a tag indicating their specific concept type delimited by square brackets:

Perhaps [*partially-justified-proposition*] she had mentioned the fact [*proposition*] already to Rebecca, but that young lady did not appear to [*partially-justified-proposition*] have remembered it [*memory-retrieval*]; indeed, vowed and protested that she expected [*add-expectation*] to see a number of Amelia's nephews and nieces. She was quite disappointed [*disappointment-emotion*] that Mr. Sedley was not married; she was sure [*justified-proposition*] Amelia had said he was, and she doted so on [*liking-emotion*] little children.

The approach that we took was to author (by hand) a set of local grammars that could be used to identify each concept. For this task we utilized the Intex Corpus Processor software developed by the Laboratoire d'Automatique Documentaire et Linguistique (LADL) of the University of Paris 7 (Silberztein, 1999). This software allowed us to author a set of local grammars using a graphical user interface, producing lexical/syntactic structures that can be compiled into finite-state transducers. To simplify the authoring of these local grammars, Intex includes a large-coverage English dictionary compiled by Blandine Courtois, allowing us to specify them at a level that generalized over noun and verb forms. For example, there are a variety of ways of expressing in English the concept of reaffirming a belief that is already held, as exemplified in the following sentences:

1) The finding was confirmed by the new data. 2) She told the truth, corroborating his story. 3) He reaffirms his love for her. 4) We need to verify the claim. 5) Make sure it is true.

Although the verbs in these sentences differ in tense, the dictionaries in Intex allowed us to recognize each using the following simple description:

(<confirm> by | <corroborate> | <reaffirm> | <verify> | <make> sure)

While constructing local grammars for each of the concepts in the original ontology of commonsense psychology, we identified several conceptual distinctions that were made in language that were not expressed in the specific concepts that Gordon had identified. For example, the original ontology included only three concepts in the conceptual area of memory retrieval (the sparsest of the 30 areas), namely *memory*, *memory cue*, and *memory retrieval*. English expressions such as toforget and repressed memory could not be easily mapped directly to one of these three concepts, which prompted us to elaborate the original sets of concepts to accommodate these and other distinctions made in language. In the case of the conceptual area of memory retrieval, a total of twelve unique concepts were necessary to achieve coverage over the distinctions evident in English.

These local grammars were authored one conceptual area at a time. At the time of the writing of this paper, our group had completed 6 of the original 30 commonsense psychology conceptual areas. The remainder of this paper focuses on the first 4 of the 6 areas that were completed, which were evaluated to determine the recall and precision performance of our hand-authored rules. These four areas are *Managing knowledge*, *Memory*, *Explanations*, and *Similarity judgments*. Figure 2 presents each of these four areas with a single fabricated example of an English expression for each of the final set of concepts. Local grammars for the two additional conceptual areas, *Goals* (20 concepts) and *Goal management* (17 concepts), were authored using the same approach as the others, but were not completed in time to be included in our performance evaluation.

After authoring these local grammars using the Intex Corpus Processor, finite-state transducers were compiled for each commonsense psychology concept in each of the different conceptual areas. To simplify the application of these transducers to text corpora and to aid in their evaluation, transducers for individual concepts were combined into a single finite state machine (one for each conceptual area). By examining the number of states and transitions in the compiled finite state graphs, some indication of their relative size can be given for the four conceptual areas that we evaluated: *Managing knowledge* (348 states / 932 transitions), *Memory* (203 / 725), *Explanations* (208 / 530), and *Similarity judgments* (121 / 500).

## 4 Performance evaluation

In order to evaluate the utility of our set of hand-authored local grammars, we conducted a study of their precision and recall performance. In order to calculate the performance levels, it was first necessary to create a test corpus that contained references to the sorts of commonsense psychological concepts that our rules were designed to recognize. To accomplish this, we administered a survey to

**1. Managing knowledge** (37 concepts)

He sgot a logical mind (*managing-knowledge-ability*). She svery gullible (*bias-toward-belief*). He sskeptical by nature (*bias-toward-disbelief*). It is the truth (*true*). That is completely false (*false*). We need to know whether it is true or false (*truth-value*). His claim was bizarre (*proposition*). I believe what you are saying (*belief*). I didn tknow about that (unknown). I used to think like you do (*revealed-incorrect-belief*). The assumption was widespread (*assumption*). There is no reason to think that (*unjustified-proposition*). There is some evidence you are right (*partially-justified-proposition*). The fact is well established (*justified-proposition*). As a rule, students are generally bright (*inference*). The conclusion could not be otherwise (*consequence*). What was the reason for your suspicion (*justification*)? That isn ta good reason (*poor-justification*). Your argument is circular (*circular-justification*). One of these things must be false (*contradiction*). His wisdom is vast (*knowledge*). He knew all about history (*knowledge-domain*). I know something about plumbing (*partial-knowledge-domain*). He sgot a lot of real-world experience (*world-knowledge*). He understands the theory behind it (*world-model-knowledge*). That is just common sense (*shared-knowledge*). I mwilling to believe that (*add-belief*). I stopped believing it after a while (*remove-belief*). I assumed you were coming (*add-assumption*). You can tmake that assumption here (*remove-assumption*). Let ssee what follows from that (*check-inferences*). Disregard the consequences of the assumption (*ignore-inference*). I tried not to think about it (*suppress-inferences*). I concluded that one of them must be wrong (*realize-contradiction*). I realized he must have been there (*realize*). I can tthink straight (*knowledge-management-failure*). It just confirms what I knew all along (*reaffirm-belief*).

**2. Memory** (12 concepts)

He has a good memory (*memory-ability*). It was one of his fondest memories (*memory-item*). He blocked out the memory of the tempestuous relationship (*repressed-memory-item*). He memorized the words of the song (*memory-storage*). She remembered the last time it rained (*memory-retrieval*). I forgot my locker combination (*memory-retrieval-failure*). He repressed the memories of his abusive father (*memory-repression*). The widow was reminded of her late husband (*reminding*). He kept the ticket stub as a memento (*memory-cue*). He intended to call his brother on his birthday (*schedule-plan*). He remembered to set the alarm before he fell asleep (*scheduled-plan-retrieval*). I forgot to take out the trash (*scheduled-plan-retrieval-failure*).

**3. Explanations** (20 concepts)

He sgood at coming up with explanations (*explanation-ability*). The cause was clear (*cause*). Nobody knew how it had happened (*mystery*). There were still some holes in his account (*explanation-criteria*). It gave us the explanation we were looking for (*explanation*). It was a plausible explanation (*candidate-explanation*). It was the best explanation I could think of (*best-candidate-explanation*). There were many contributing factors (*factor*). I came up with an explanation (*explain*). Let sfigure out why it was so (*attempt-to-explain*). He came up with a reasonable explanation (*generate-candidate-explanation*). We need to consider all of the possible explanations (*assess-candidate-explanations*). That is the explanation he went with (*adopt-explanation*). We failed to come up with an explanation (*explanation-failure*). I can tthink of anything that could have caused it (*explanation-generation-failure*). None of these explanations account for the facts (*explanation-satisfaction-failure*). Your account must be wrong (*unsatisfying-explanation*). I prefer non-religious explanations (explanation-preference). You should always look for scientific explanations (*add-explanation-preference*). We renot going to look at all possible explanations (*remove-explanation-preference*).

**4. Similarity judgments** (13 concepts)

She sgood at picking out things that are different (*similarity-comparison-ability*). Look at the similarities between the two (*make-comparison*). He saw that they were the same at an abstract level (*draw-analogy*). She could see the pattern unfolding (*find-pattern*). It depends on what basis you use for comparison (*comparison-metric*). They have that in common (*same-characteristic*). They differ in that regard (*different-characteristic*). If a tree were a person, its leaves would correspond to fingers (*analogical-mapping*). The pattern in the rug was intricate (*pattern*). They are very much alike (*similar*). It is completely different (*dissimilar*). It was an analogous example (*analogous*).

Figure 2. Example sentences referring to 92 concepts in 4 areas of commonsense psychology

collect novel sentences that could be used for this purpose.

This survey was administered over the course of one day to anonymous adult volunteers who stopped by a table that we had set up on our university scampus. We instructed the survey taker to author 3 sentences that included words or phrases related to a given concept, and 3 sentences that they felt did not contain any such references. Each survey taker was asked to generate these 6 sentences for each of the 4 concept areas that we were evaluating, described on the survey in the following manner:

> *Managing knowledge*: Anything about the knowledge, assumptions, or beliefs that people have in their mind
> *Memory*: When people remember things, forget things, or are reminded of things
> *Explanations*: When people come up with possible explanations for unknown causes
> *Similarity judgments*: When people find similarities or differences in things

A total of 99 people volunteered to take our survey, resulting in a corpus of 297 positive and 297 negative sentences for each conceptual area, with a few exceptions due to incomplete surveys.

Using this survey data, we calculated the precision and recall performance of our hand-authored local grammars. Every sentence that had at least one concept detected for the corresponding concept area was treated as a hit .Table 1 presents the precision and recall performance for each concept area. The results show that the precision of our system is very high, with marginal recall performance.

The low recall scores raised a concern over the quality of our test data. In reviewing the sentences that were collected, it was apparent that some survey participants were not able to complete the task as we had specified. To improve the validity of the test data, we enlisted six volunteers (native English speakers not members of our development team) to judge whether or not each sentence in the corpus was produced according to the instructions. The corpus of sentences was divided evenly among these six raters, and each sentence that the rater judged as not satisfying the instructions was filtered from the data set. In addition, each rater also judged half of the sentences given to a different rater in order to compute the degree of inter-rater agreement for this filtering task. After filtering sentences from the corpus, a second precision/recall evaluation was performed. Table 2 presents the results of our hand-authored local grammars on the filtered data set, and lists the inter-rater agreement for each conceptual area among our six raters. The results show that the system achieves a high level of precision, and the recall performance is much better than earlier indicated.

The performance of our hand-authored local grammars was then compared to the performance that could be obtained using more traditional machine-learning approaches. In these comparisons, the recognition of commonsense psychology concepts was treated as a classification problem, where the task was to distinguish between positive

| Concept area | Correct Hits (a) | Wrong hits (b) | Total positive sentences (c) | Total negative sentences | Precision (a/(a+b)) | Recall (a/c) |
|---|---|---|---|---|---|---|
| Managing knowledge | 205 | 16 | 297 | 297 | 92.76% | 69.02% |
| Memory | 240 | 4 | 297 | 297 | 98.36% | 80.80% |
| Explanations | 126 | 7 | 296 | 296 | 94.73% | 42.56% |
| Similarity judgments | 178 | 18 | 296 | 297 | 90.81% | 60.13% |
| | 749 | 45 | 1186 | 1187 | 94.33% | 63.15% |

Table 1. Precision and recall results on the unfiltered data set

| Concept area | Inter-rater agreement (*K*) | Correct Hits (a) | Wrong hits (b) | Total positive sentences (c) | Total negative sentences | Precision (a/(a+b)) | Recall (a/c) |
|---|---|---|---|---|---|---|---|
| Managing knowledge | 0.5636 | 141 | 12 | 168 | 259 | 92.15% | 83.92% |
| Memory | 0.8069 | 209 | 0 | 221 | 290 | 100% | 94.57% |
| Explanations | 0.7138 | 83 | 5 | 120 | 290 | 94.21% | 69.16% |
| Similarity judgments | 0.6551 | 136 | 12 | 189 | 284 | 91.89% | 71.95% |
| | 0.6805 | 569 | 29 | 698 | 1123 | 95.15% | 81.51% |

Table 2. Precision and recall results on the filtered data set, with inter-rater agreement on filtering

| | A. Hand authored local grammars | | B. SVM with word level features | | C. SVM with word and concept features | |
|---|---|---|---|---|---|---|
| Concept area | *Pa* | *K* | *Pa* | *K* | *Pa* | *K* |
| Managing knowledge | 90.86% | 0.8148 | 86.0789% | 0.6974 | 89.5592% | 0.7757 |
| Memory | 97.65% | 0.8973 | 93.5922% | 0.8678 | 97.4757% | 0.9483 |
| Explanations | 89.75% | 0.7027 | 85.9564% | 0.6212 | 89.3462% | 0.7186 |
| Similarity judgments | 86.25% | 0.7706 | 92.4528% | 0.8409 | 92.0335% | 0.8309 |

Table 3. Percent agreement (*Pa*) and Kappa statistics (*K*) for classification using hand-authored local grammars (A), SVMs with word features (B), and SVMs with word and concept features (C)

and negative sentences for any given concept area. Sentences in the filtered data sets were used as training instances, and feature vectors for each sentence were composed of word-level unigram and bi-gram features, using no stop-lists and by ignoring punctuation and case. By using a toolkit of machine learning algorithms (Witten & Frank, 1999), we were able to compare the performance of a wide range of different techniques, including Na□veBayes, C4.5 rule induction, and Support Vector Machines, through stratified cross-validation (10-fold) of the training data. The highest performance levels were achieved using a sequential minimal optimization algorithm for training a support vector classifier using polynomial kernels (Platt, 1998). These performance results are presented in Table 3. The percentage correctness of classification (*Pa*) of our hand-authored local grammars (column A) was higher than could be attained using this machine-learning approach (column B) in three out of the four concept areas.

We then conducted an additional study to determine if the two approaches (hand-authored local grammars and machine learning) could be complimentary. The concepts that are recognized by our hand-authored rules could be conceived as additional bimodal features for use in machine learning algorithms. We constructed an additional set of support vector machine classifiers trained on the filtered data set that included these additional concept-level features in the feature vector of each instance along side the existing unigram and bi-gram features. Performance of these enhanced classifiers, also obtained through stratified cross-validation (10-fold), are also reported in Table 3 as well (column C). The results show that these enhanced classifiers perform at a level that is the greater of that of each independent approach.

## 5 Discussion

The most significant challenge facing developers of large-scale lexical-semantic resources is coming to some agreement on the way that natural language can be mapped onto specific concepts. This challenge is particularly evident in consideration of our survey data and subsequent filtering. The abilities that people have in producing and recognizing sentences containing related words or phrases differed significantly across concept areas. While raters could agree on what constitutes a sentence containing an expression about memory (Kappa=.8069), the agreement on expressions of managing knowledge is much lower than we would hope for (Kappa=.5636). We would expect much greater inter-rater agreement if we had trained our six raters for the filtering task, that is, described exactly which concepts we were looking for and gave them examples of how these concepts can be realized in English text. However, this approach would have invalidated our performance results on the filtered data set, as the task of the raters would be biased toward identifying examples that our system would likely perform well on rather than identifying references to concepts of commonsense psychology.

Our inter-rater agreement concern is indicative of a larger problem in the construction of large-scale lexical-semantic resources. The deeper we delve into the meaning of natural language, the less we are likely to find strong agreement among untrained people concerning the particular concepts that are expressed in any given text. Even with lexical-semantic resources about commonsense knowledge (e.g. commonsense psychology), finer distinctions in meaning will require the efforts of trained knowledge engineers to successfully map between language and concepts. While this will certainly create a problem for future preci-

sion/recall performance evaluations, the concern is even more serious for other methodologies that rely on large amounts of hand-tagged text data to create the recognition rules in the first place. We expect that this problem will become more evident as projects using algorithms to induce local grammars from manually-tagged corpora, such as the Berkeley FrameNet efforts (Baker et al., 1998), broaden and deepen their encodings in conceptual areas that are more abstract (e.g. commonsense psychology).

The approach that we have taken in our research does not offer a solution to the growing problem of evaluating lexical-semantic resources. However, by hand-authoring local grammars for specific concepts rather than inducing them from tagged text, we have demonstrated a successful methodology for creating lexical-semantic resources with a high degree of conceptual breadth and depth. By employing linguistic and knowledge engineering skills in a combined manner we have been able to make strong ontological commitments about the meaning of an important portion of the English language. We have demonstrated that the precision and recall performance of this approach is high, achieving classification performance greater than that of standard machine-learning techniques. Furthermore, we have shown that hand-authored local grammars can be used to identify concepts that can be easily combined with word-level features (e.g. unigrams, bi-grams) for integration into statistical natural language processing systems. Our early exploration of the application of this work for corpus analysis (U.S. State of the Union Addresses) has produced interesting results, and we expect that the continued development of this resource will be important to the success of future corpus analysis and human-computer interaction projects.

## Acknowledgments

This paper was developed in part with funds from the U.S. Army Research Institute for the Behavioral and Social Sciences under ARO contract number DAAD 19-99-D-0046. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Army.

## References

Baker, C., Fillmore, C., & Lowe, J. (1998) The Berkeley FrameNet project. in *Proceedings of the COLING-ACL*, Montreal, Canada.

Bartsch, K. & Wellman, H. (1995) *Children talk about the mind.* New York: Oxford University Press.

Dyer, J., Shatz, M., & Wellman, H. (2000) Young children sstorybooks as a source of mental state information. *Cognitive Development 15*:17-37.

Ferguson, G. & Allen, J. (1993) Cooperative Plan Reasoning for Dialogue Systems, in *AAAI-93 Fall Symposium on Human-Computer Collaboration: Reconciling Theory, Synthesizing Practice, AAAI Technical Report FS-93-05*. Menlo Park, CA: AAAI Press.

Gordon, A. (2002) The Theory of Mind in Strategy Representations. *24th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gordon, A. & Hobbs (2003) Coverage and competency in formal theories: A commonsense theory of memory. AAAI Spring Symposium on Formal Theories of Commonsense knowledge, March 24-26, Stanford.

Platt, J. (1998). *Fast Training of Support Vector Machines using Sequential Minimal Optimization.* In B. Sch□lkopf,C. Burges, and A. Smola (eds.) Ad*vances in Kernel Methods - Support Vector Learning*, Cambridge, MA: MIT Press.

Reboul A., Sabatier P., No□lJorand M-C. (2001) Le discours des schizophr□nes:une □tudede cas. *Revue fran□aisede Psychiatrie et de Psychologie M□dicale*, *49*, pp 6-11.

Silberztein, M. (1999) Text Indexing with INTEX. *Computers and the Humanities* 33(3).

Traum, D. (1993) Mental state in the TRAINS-92 dialogue manager. In *Working Notes of the AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, pages 143-149, 1993. Menlo Park, CA: AAAI Press.

Voorhees, E. & Buckland, L. (2002) *The Eleventh Text REtrieval Conference (TREC 2002).* Washington, DC: Department of Commerce, National Institute of Standards and Technology.

Witten, I. & Frank, E. (1999) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufman.