Acquiring Commonsense Knowledge for a Cognitive Agent
James Allen,1,2 William de Beaumont,2 Nate Blaylock,2 George Ferguson,1
Jansen Orfan,1 Mary Swift1
1University of Rochester
2Institute for Human and Machine Cognition
{james, ferguson, jorfan, swift}@cs.rochester.edu
{blaylock, wbeaumont}@ihmc.us

Abstract
A critical prerequisite for human-level cognitive systems is having a rich conceptual understanding of the world. We describe a system that learns conceptual knowledge by deep understanding of WordNet glosses. While WordNet is often criticized for having a too fine-grained approach to word senses, the set of glosses do generally capture useful knowledge about the world and encode a substantial knowledge base about everyday concepts. Unlike previous approaches that have built ontologies of atomic concepts from the provided WordNet hierarchies, we construct complex concepts compositionally using description logic and perform reasoning to derive the best classification of knowledge. We view this work as simultaneously accomplishing two goals: building a rich semantic lexicon useful for natural language processing, and building a knowledge base that encodes common-sense knowledge.

Introduction
When researchers talk about learning by reading, they often are talking about different types of knowledge that is acquired. To us, there are three main classes of knowledge that need to be acquired to enable a human-level cognitive agent:

1. Learning facts about the world (e.g., Chicago is a city, Chicago is in the USA,...)
2. Learning complex structured information using existing concepts (e.g., how to get to a restaurant)
3. Learning new concepts that extend our ontology (e.g., learning about plumbing)

This paper describes work addressing the third problem: how to acquire new concepts and learn the meanings of new words. It is motivated by our long-standing interest in building deep language understanding systems that provide a high-fidelity mapping of language to an underlying knowledge representation (KR). Rather than trying to build the entire target ontology by hand, however, we want to build the ontology and its associated common-sense knowledge by reading existing textual sources. We believe this is the only way we may be able to build truly broad-coverage deep understanding systems, which then would enable the acquisitions of much more complex common-sense knowledge by reading further texts (e.g., Wikipedia).

As an initial exploration of the feasibility of this enterprise, we are building an ontology by reading the definitions that are present in WordNet glosses. Learning concepts from machine readable dictionaries has a long history. Ide & Veronis (1994) is an excellent survey of early work, and lays out key problems in understanding dictionary definitions. First, definitions tend to be quite complex, both in structure and in the conceptualizations used; and second, they are vague, imprecise, and incomplete. However, while glosses may not capture word meaning precisely, they can capture significant amounts of common-sense knowledge.

We believe we can overcome the difficulties others have found by bringing several key ideas to bear. First, we bootstrap the whole process with a hand-built upper-level ontology that works in conjunction with a deep understanding system, namely the TRIPS system (Allen et al, 2008). This provides a capability to interpret many of the complexities in the glosses, and seeds the development of the ontology inherent in the glosses. Second, we define concepts compositionally using description logic (specifically OWL-DL) and then use off-the-shelf reasoning systems to develop the ontology using efficient subsumption algorithms. The result is a multiple-inheritance structure supporting complex composition concepts, including disjunctive and conjunctive concepts, much better suited to the definitions present in WordNet (Miller, 1995). Third, in order to produce a useful common-sense knowledge base, we hand-axiomatize key abstract concepts that people would learn from direct experience with the world, and are difficult to capture in language definitions. These are concepts like START, STOP, CHANGE, and the like. With fairly rudimentary axiomatization of such core concepts, the learned knowledge is “activated” and can produce significant entailments and common-sense models from the data. And finally, though not discussed in this paper, we use similarity and abstraction techniques to clean up the conceptual hierarchy, merging redundant and/or overly specific concepts, producing an ontology better suited for both reasoning and language processing.

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

A major stumbling block in learning conceptual knowledge from text has been the differences between the semantics inherent in natural language and the knowledge representations that have been developed for reasoning. One key problem is the issue of granularity. Linguistic representations primarily encode meaning at the word level, whereas knowledge representations deal with more complex conceptualization. As an example, note that KRs often have a top-level distinction that subdivides events into processes (perdurants in DOLCE (Gangemi et al. 2002)) such as I’m laughing and accomplishments (endurants in DOLCE) such as It broke. But when we try to define the meaning of the verb climb, where does it fit? The sentence I climbed all day describes a process, whereas I climbed the hill describes an accomplishment. Is the verb climb ambiguous? If not, how can we classify it? The problem is that many concepts in ontologies correspond to phrasal meanings rather than word meanings, whereas linguistic ontologies stay focused on the words. This is one reason why WordNet has so many senses for each word, it tries to pack meanings of phrases into meanings of single words (typically the verb).

To unify the two styles of representation: we must ground out the representation in word meanings, because that is how the knowledge in language is delivered, and then capture entailments using complex concepts built compositionally from word meanings. With this approach, a verb like climb can have a single sense and the events that correspond to sentences correspond to complex concepts that involve this sense.

Specifically, we are employing the deep understanding capabilities of the TRIPS system, with its extensive grammar, lexicon and 2000+ concept upper ontology, to parse WordNet glosses. Like many other knowledge representation systems such as Cyc (Lenat & Guha 1990), KM (Clark & Porter 1996) and SUMO (Niles & Pease 2001), TRIPS includes a coarse grained mapping from WordNet synsets into its upper ontology. This allows an initial scaffolding with which to construct the much more detailed ontology based on the glosses.

The resulting knowledge is represented in OWL (Smith et al. 2004), for which there is a substantial technology base of efficient reasoning engines. Key to our effort, we use subsumption reasoning to organize and refine new concepts into a rich ontology. Also key is the ability of an OWL-based system to support complex compositional concepts, not just atomic concepts. This allows us to construct complete representations of many definitions, and reasonable approximations for the more complex definitions. Note that while many criticize WordNet for having too fine-grained a treatment of word senses, it’s important to note that the glosses in WordNet are generally true statements about word/concept meaning, and thus encode common-sense knowledge as reflected in word meaning. By reasoning over the definitions of the different variants for a word, we can develop techniques to collapse similar senses/synsets in a systematic way.

Consider a simple example of our approach: the WordNet synset wake_up%2-29-00, involving the verb “wake up,” is defined by the gloss stop sleeping[1:09:00]. Our processing is made simpler by the fact that about half the content words in WordNet glosses are already tagged with their WordNet sense. In this definition, we are given the sense of sleeping but must identify the appropriate sense of stop.

The relevant part of the logical form produced by the TRIPS parser for this definition is output as three terms: (F v1 (* STOP stop%2:42:00) :affected v3 :process v2) (F v2 (* BODILY-PROCESS sleep%1:09:00) :affected v3) (IMPRO v3 LIVING)

Interpreting these terms, there is an event, v1, which is of type STOP in the TRIPS ontology and stop%2:38:00 in WordNet, which has a :process role filled with an event v2, which is of type BODILY-PROCESS in TRIPS and sleep%1:09:00 in Wordnet. Finally, there is an implicit argument (the IMPRO form), which is an object of type LIVING and fills the :affected role of both the stopping and sleeping events. This type was derived from the selectional restrictions in the TRIPS ontology that only LIVING things can undergo bodily processes.

The TRIPS ontology gives an initial framework for conceptual classification but few specifics about actual word meaning for most words. For instance, TRIPS knows nothing about the action sleeping%1-09-00, except that it is some BODILY-PROCESS. Since we haven’t read the definition of sleeping%1-09-00 yet, we have no further information about it.

The other key thing we need to do to define a new concept is to identify the its arguments (i.e., roles). To identify the roles for new concepts, we take all the roles in the defining concept that are either missing in the definition (i.e., as a gap in the definition) or are filled by an indefinite pro-form (e.g., someone, somewhere, ...). In the current example, we infer that the new concept “wake up” has an affected role, derived from the role structure of STOP. Note another essential role in STOP, namely the :process role, but this is not inherited to wake_up%2-29-00 because it was fully instantiated in the definition. To confirm the role hypotheses, we parse the examples attached to entries in wordnet and adjust the analysis if necessary.
Figure 1 shows the definition in human-readable form that the system constructs from the gloss. The double arrows relate an entire concept to its superclass, and the single arrows relate an argument to its superclass (i.e., its restriction). The actual definition in OWL-DL would be:

\[ \text{wake up}^{2:29:00} \subseteq \text{stop}^{2:42:00} \cap \exists \text{process sleeping}^{1:09:00} \cap \exists \text{affected living} \]

i.e., a waking up event is a subset of stopping events whose \text{affected} role is a living being and whose \text{process} is sleeping.

When we later read the gloss for \text{sleeping}^{1:09:00} we can extend the knowledge base with additional information. The definition of \text{sleeping}^{1:09:00} is the \text{state}^{1:03:00} of being \text{asleep}^{4:02:00}, and the result of processing this is shown in Figure 2.

As mentioned above, a key part of our approach is that we treat concepts as complex objects, not atoms. This allows for compositional construction of new concepts as well as opening the door to using reasoning techniques such as the subsumption algorithm found in description logics. An ontology that treats all concepts as atoms inhibits learning, for there is no systematic mapping from phrases in language to new concepts. There is no structure to support non-trivial reasoning.

Before we consider how we use such knowledge for reasoning, we will explore the construction of concepts from definitions in more detail.

**Constructing Definitions**

This section provides a little more detail on how we build definitions from glosses. The first stage of processing involves parsing the definitions using TRIPS parser [Allen et al, 2008]. We use all the information present in the glosses, including part of speech tags and sense tags when available, in order to provide guidance to the parser. For words not in the TRIPS lexicon, or that do not have a sense corresponding to the WordNet sense tag, the TRIPS lexicon managers searches WordNet for words matching the provided information and automatically constructs an underspecified representation in the TRIPS lexical format. The WordNet sense tags are converted into TRIPS ontology concepts using an existing hand-built mapping from high-level WordNet senses to the TRIPS ontology. To control ambiguity arising from multiple senses in WordNet, the parser prefers lexical entries corresponding to actual tagged senses when they are present in the glosses.

<table>
<thead>
<tr>
<th>Synset Name</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>wake up^{2:29:00}</td>
<td>to stop^{2:42:00} sleeping^{1:09:00}</td>
</tr>
<tr>
<td>sleeping^{1:09:00}</td>
<td>the state^{1:03:00} of being asleep^{4:02:00}</td>
</tr>
<tr>
<td>asleep^{4:02:00}</td>
<td>into a sleeping state^{1:03:00}</td>
</tr>
<tr>
<td>fall asleep^{2:29:00}</td>
<td>to change^{2:30:00} from a waking^{1:09:00} to a sleeping [sleep^{2:29:00}] state^{1:03:00}</td>
</tr>
<tr>
<td>waking^{1:09:00}</td>
<td>the state^{1:03:00} of remaining awake</td>
</tr>
<tr>
<td>sleep^{2:29:00}</td>
<td>to be asleep^{4:02:00}</td>
</tr>
<tr>
<td>awake^{3:00:00}</td>
<td>not in a state^{1:03:00} of sleep^{1:26:00}</td>
</tr>
<tr>
<td>asleep^{3:00:00}</td>
<td>in a state^{1:03:00} of sleep^{1:26:00}</td>
</tr>
<tr>
<td>sleep^{1:26:00}</td>
<td>a natural and periodic state^{1:03:00} of rest during which consciousness^{1:09:00} of the world^{1:09:00} is suspended [suspend^{2:30:02}]</td>
</tr>
<tr>
<td>keep up^{2:29:00}</td>
<td>prevent^{2:41:00} from going to bed [go to bed^{2:29:00}]</td>
</tr>
<tr>
<td>go to bed^{2:29:00}</td>
<td>to prepare for sleep^{2:29:00}</td>
</tr>
<tr>
<td>sleeper^{1:18:00}</td>
<td>a rester^{1:18:00} who is sleeping [sleep^{2:29:00}]</td>
</tr>
<tr>
<td>rester^{1:18:00}</td>
<td>a person^{1:03:00} who rests [rest^{2:35:00},rest^{2:32:00},rest^{2:29:00}]</td>
</tr>
</tbody>
</table>

*Table 1: Some Glosses in WordNet directly related to sleeping, showing sense tags when provided*
The output of the parser is a logical form graph, which encodes an unscoped modal logic with reified events (Manshadi et al, 2008) and semantic roles as shown above.

The logical form graph is then translated into an OWL compositional class using a simple algorithm, which we illustrate using the LF term

1. We create a new concept for the class being defined using the type in the LF graph (i.e., \textit{wake\_up}\%2:29:00). Call this $C$.

2. Create (or identify in the ontology) a concept corresponding to the defining concept (i.e., \textit{stop}\%2:42:00). Call this $T$.

3. If the defining concept has roles $r_1, \ldots, r_n$ (e.g., \texttt{:affected}, \texttt{:process}), recursively build new concepts for each of the arguments. Call these concepts $R_1, \ldots, R_n$.

4. Define the class: $C \subseteq T \cap \exists r_1 R_1 \cap \ldots \cap \exists r_n R_n$

There are a few special cases that complicate the algorithm. For example, many of the glosses involve disjunctive constructs, and these are handled by mapping the disjunctive in the logical form to a \texttt{UNION} of the two classes. Conjunctions are treated as intersections.

**Building a Micro-Theory of Sleeping**

To explore the potential of this approach, we extracted a subset of WordNet consisting of all the concepts that involve the notion of sleeping in some way. This includes all the senses of sleep itself, words derivationally related to sleep, words that have some connection to sleep in their definitions, plus the definitions of all words used in the definitions of sleep words (see Table 1). WordNet does encode some relationships between these words, but they are highly underspecified. The most common link is the “derivationally related” link, which indicates some morphological connection between words but nothing about their semantic relationship. More useful are the antonym relations, and we use these links when gathering related words. We did not, for this experiment, use the antonym information to define the KB. Thus one test of the worth of our approach is whether antonyms can be derived automatically.

The result of interpreting these glosses is shown in Figure 3. There are few things to note. First, while there is considerable variation in how definitions are phrased, in the end all the sense reduce down to the nominal concept \texttt{sleep}\%1:26:00. This concept is further defined (as a natural and periodic state of rest during which consciousness of the world is suspended), but we have suppressed this information as it does not affect our example. On the face of it, there seems great potential for reasoning over these concepts. We see, for instance, that being \texttt{awake}\%3:00:00 is “not being in state of \texttt{sleep}\%1:26:00” while being \texttt{asleep}\%3:00:00 is “in a state of \texttt{sleep}\%1:26:00”. With an appropriate handling of negation, we could infer that both these states cannot hold simultaneously. But other conclusions seem close but evade realization. For instance, we would like to be able to conclude that after one wakes up, one is awake. But this not not derivable from the current knowledge base as we
don’t have appropriate knowledge about what it means to stop%2:42:00. And looking at the WordNet definitions doesn’t help. Here are the relevant definitions:

Stop: come to a halt, stop moving
Halt: the event of something ending
End: the point in time when something ends

WordNet has the same problem that we find in all dictionaries, namely that some core concepts are simply impossible to reduce to other concepts and the definitions become circular. This leads us to the next key part of the work: defining an inferential model that can exploit the knowledge we have gained.

Inference

Description logic supports inference about classes of objects, and provides efficient inference models for computing subsumption of classes in a multiple-inheritance representation. These definitions say nothing about any particular individuals. For instance, wake_up%2:29:00 was defined as the stopping of an event described by the nominalization sleep%1:09:00, which is defined as “a state of sleep%2:29:00”, which is the verbal form of the event. Nothing in the description logic says that when you are in the state of some event occurring, that the event is actually occurring. To capture such relationships and use the knowledge we produce, we need to introduce some axioms for basic concepts.

We’ll define a representation for making assertions based on interval-based temporal logic (Allen, 1984). We have a predicate T that asserts that some concept (described by a OWL concept) holds/occurs over time t. While Allen distinguished between events occurring and properties holding, we will ignore such issues here as they do not affect the examples. With this representation, we can now define some basic axioms that enable inference. To handle negation, we have the following axiom (where all non-quantified lower case letters are universally quantified variables)

\[ T(\text{NOT :arg p}, t) \leftrightarrow \neg T(p, t) \]

i.e., the concept of a situation not holding holds at t if and only if it’s not the case that the situation holds at t. With this axiom, we can now prove that the situations described by awake%3:00:00 and asleep%3:00:00 are mutually exclusive.

In order to derive many of the other conclusions we’d like, we need to axiomatize some key concepts like stopping, causing, preventing, and so on. As an example, here is one of the axioms for stop%2:42:00:

\[ T[\text{stop%2:42:00 :process e}, t] \supset \exists t' \cdot \neg T(e, t') \land T(e, t) \land \neg T[\text{process e}, t'] \land \neg T[\text{process e}, t] \land T[e, t] \land T[e, t'] \]

i.e., if process e stops at time t, then e is not occurring at time t’ that immediately follows t. The meets relation means one time immediately follows another and is defined in Allen (1983).

Figure 4 shows a sampling of these core axioms, ones that are particularly relevant to the micro-theory of sleeping. A reader might complain that we are hand-building the knowledge we need, but in fact, we believe the number of core concepts that will need to be axiomatized is relatively small, say on the order of 50 to 100 concepts, and that almost all the knowledge in the axioms is knowledge that a cognitive agent would generally learn by early experiences interacting with the world. So the hand-defined set of axioms replace the the need for an agent that can learn in situ before acquiring much language. With this base, then we can learn the meanings of all other words directly from the glosses.

Finally, in this logical framework we need an axiom schema that relates subsumption and inference. Essentially, if a concept describing a situation holds, then all superconcepts of that concept also hold. This rule is the inheritance rule:

Inheritance Rule

If concept A subsumes concept B, then \( T(B, t) \supset T(A, t) \)

We now have everything we need to use the knowledge base to compute entailments. First, note we can derive the antonym relation between awake%3:00:00 and asleep%3:00:00 directly from the definitions.

As a more complex example, consider the following proof of the statement Before falling asleep (Fall_asleep%2:29:00), the agent is awake%3:00:00:

a. Assume \( T(\text{Fall_asleep%2:29:00}, t) \)
b. Thus [change%2:30:02 :to [state%1:03:00 :of sleep%2:29:00]] holds at time T (using defn of Fall_asleep%2:29:00)

c. sleep%1:09:00 holds at a time Tn which immediately follows time T (axiom 3)

d. [state%1:03:00 :of [HAVE-PROPERTY :property asleep%3:00:00]] holds at the a time Tn ([defn of sleep%1:09:00])

e. [HAVE-PROPERTY :property asleep%3:00:00] holds at Tn (axiom 8).

f. asleep%3:00:00 holds at Tn (axiom 4)

g. [in%4:02:01 :val [state%1:03:00 :of sleep%1:26:00]] holds at time Tn (defn, asleep%3:00:00)

h. [state%1:03:00 :of sleep%1:26:00] holds at time Tn (axiom 2)

i. Thus, sleep%1:26:00 holds at time Tn (axiom 3)

Note that while there are 8 inference steps to derive this conclusion from the definitions, the search space is very small and such inference can be performed rapidly using a simple forward inference chaining algorithm. It would be entirely possible to precompute many of these conclusions (that seem obvious to us) in a post-processing phase of the learning in order to simplify the knowledge base.

Implementation

We have implemented this reasoning framework using SWRL, a standard reasoning engine for OWL. Each of the axioms in Figure 4 have a fairly direct translation into SWRL. We then perform inference in a forward chaining fashion when instances are added, pretty much following the proof just presented.

Using this, we can test what conclusions the agent can draw about sleeping given the derived knowledge from the glosses. Here are a few other statement it now believes

Being awake%3:00:0 is mutually exclusive with being asleep%3:00:00

When an agent is Sleeping%2:29:00, then it is asleep, and thus not awake.

When an agent falls asleep (Fall_asleep%2:29:00), then it is in a state of sleep afterwards, and is not awake.

If an agent wakes up, then it was sleeping and now is awake.

Note that there are other facts we would like to derive but cannot be obtained from the definitions. For instance, we learn that keeping someone up is preventing them from going to bed, and going-to-bed is defined as preparing for sleep. So we can conclude that keeping someone up prevents them from preparing for sleep, but not that they didn’t fall asleep anyway without preparation! Either we need to find a better definition or we need to develop a framework for default/probabilistic reasoning. Probably the most practical solution is to gather information from many different sources and use reasoning processes to combine and reconcile differences.

Can this really work?

Constructing the mini-theory of sleep was quite successful, so it leads to the question of whether we can use these techniques to construct a knowledge base containing all the information present in WordNet. In this section we discuss some of the challenges we face in accomplishing this task. There are two types of challenges we face: getting accurately understanding of definitions, and cleaning up and organizing the knowledge that is produced.

Challenges in Understanding

The glosses for the sleeping words made for a good example because the definitions are unusually clean and concise. Many other glosses in WordNet are significantly more difficult to understand, which we discuss here.

Disjunction and Conjunction

Many glosses involve significant use of disjunctions and conjunctions. While it is not difficult to express such concepts in OWL-DL, getting the right parses is a considerable challenge. Consider the gloss for bouillabaisse:

Highly seasoned Mediterranean soup or stew made of several kinds of fish and shellfish with tomatoes and onions or leeks and seasoned with saffron and garlic and herbs

The TRIPS parser is unable to find a full spanning parse for this definition because of the complexity of the search space. It does however produce a partial representation using its robustness mechanism that is equivalent to the fragment a seasoned Mediterranean soup or stew. While this correctly defines bouillabaisse as a soup/stew dish it misses much of the detail present in the gloss. We plan to explore some preprocessing techniques to bracket conjunctions using semantic similarity measures to improve conjunction handling.

Vague Semantic Relations

We can use the same example to illustrate the next issue. The system is currently unable to refine vague relations in the definitions. The most common source of such vagueness is in noun-noun modification, as in Mediterranean soup. The logical form for this fragment uses the very abstract relation Assoc-with (i.e., associated with), which could be ownership, part-of, made-from, and so on. We currently have no mechanism for refining this to a better relation. This may be a case where we attempt a bootstrapping approach. We first acquire information about the various relationships such as partof and madeof using text mining techniques. We then can use this knowledge to identify more specific relations when available. As an example, the gloss for the locomotive sense of engine starts with a wheeled vehicle ... which produces an LF with a
vehicle \textit{Assoc-with} wheels. If we already know that vehicles can have wheels, we could refine this semantic relation appropriately. Note that while WordNet does contain \textit{part-of} information, e.g., there is only one part identified for vehicles, namely the dashboard.

\textbf{Word Sense Ambiguity}

In general word sense ambiguity remains an issue when learning by reading online sources. For WordNet, however, about half of the words in the the glosses have been hand-tagged with their senses, making this a very useful resource to start with. While there are some errors, as noted above, in some cases the parser is able to correct the problem as these tagged senses only come in as preferences, not absolute constraints. For addressing the general problem when we move to other resources, we need to first identify what are useful senses. It is widely acknowledged that WordNet uses far too fine a set of senses to be useful as a knowledge base. This issue is revisited in the next section, under cleaning up and organizing knowledge.

\textbf{Challenges in Organizing Acquired Knowledge}

Even if the glosses were understood perfectly, the resulting knowledge would be a jumble of idiosyncratic information based on the whims of the people writing the glosses. In this section we consider a few techniques that we pursuing to clean up the resulting knowledge.

\textbf{Cleaning up Redundancies}

Note first that we do have a good start in using OWL-DL reasoners. If equivalent concepts are defined from the glosses, the subsumption algorithms will be able to identify them. There are a fair number of concepts in WordNet that are unnecessary and create spurious ambiguity issues. For instance, WordNet has a concept \textit{young_person}\%1:18:00 defined as \textit{a young person!} To identify such issues, we can parse the phrase \textit{young person} that is associated with the sense and find that the compositional semantics is exactly equivalent to the gloss definition. In other words, there is no need for \textit{young_person}\%1:18:00 as a separate concept as it can be produced compositionally. Not all cases are so easy, but we believe similar techniques will identify likely candidates. For instance, \textit{young_woman}\%1:18:00 is defined as \textit{a young female}, and \textit{woman}\%1:18:01 is defined as \textit{a female person}. Processing these definitions places \textit{young_woman}\%1:18:00 very close in the hierarchy to the compositional meaning of \textit{young woman}. We need to identify heuristics for collapsing such cases.

\textbf{Reducing Word Senses}

Related to this are techniques we are exploring for reducing the number of senses for words to an effective core. As an example, consider glosses for two of the 49 senses of the verb \textit{make}:

\textit{Cause to do}

\textit{Compel or make someone to act in a certain way}

Now, \textit{compel} is defined as \textit{force somebody to do something}, and \textit{force} is defined as \textit{to cause to do}. Thus \textit{compel} and \textit{cause} are closely related. Furthermore, the appropriate sense of \textit{do} is defined as \textit{perform an action}, and \textit{act} is also defined as \textit{perform an action}. By constructing all these concepts, we find these two senses of \textit{make} are virtual synonyms of each other, and would construct a single sense that captures both.

This, again, is a simple example. We are exploring more general clustering techniques based on semantic similarity in the conceptual hierarchy to identify abstract senses that produce much cleaner knowledge.

\textbf{Eliminating Meta-talk and Language Variants}

Language supports substantial variation in conceptualizing the same circumstances. This can be seen in the sleep mini-theory we constructed, where we see multiple nominalizations (\textit{sleeping\%1:09:00}, \textit{sleep\%1:26:00}), verbal forms (\textit{sleep\%1:26:00}), adjectival properties (\textit{asleep\%3:00:00}) and adverbials (\textit{asleep\%4:02:00}) all essentially referring to the same state of affairs (i.e., being asleep). One of the points of doing the sleeping mini-theory in detail was to show that we could inferentially connect all these concepts in a reasonable way. We could also use the same derived knowledge, however, to produce much more concise knowledge base that is abstracted one level away from language. The idea would be to use the forward chaining of the core axioms in order to produce simplified knowledge. As an example, consider the concept \textit{wake_up\%2:29:00} again. Rather than depending on inference every time we consider a waking up event, we could do forward chaining once in a preprocessing stage and then build a richer conceptual structure with the concept. Taking terminology from the event semantics, we might directly assert that the result of a \textit{wake up} event is that the actor is \textit{awake}, and that a necessary precondition on this event occurring is that the actor was \textit{asleep}. By doing this, we could build an axiomatization of events similar to that proposed in Allen (1984).

\textbf{Related Work}

While clearly a critical need in constructing human-level cognitive agents, automatic acquisition of linguistically-based knowledge has not be explored much in the cognitive systems literature. But there is a significant history trying to build lexical knowledge from dictionaries, and a substantial amount of new work on mining the web to learn facts. The latter work, however, does not focus on learning conceptual knowledge, so we will not consider it further here.

Research on automatically building knowledge bases from lexical resources has a long history. Early attempts used techniques such as hypernym extraction (e.g., Alshawi 1989; Chodorow et al. 1985), pattern matching (e.g., Alshawi 1989; Vossen et al. 1989; Wilks et al. 1989) and co-occurrence data (e.g., Wilks et al. 1989).
In an evaluation, Ide & Véronis (1994) conclude that the limited results from these efforts show that understanding lexical definitions requires deeper processing than the techniques applied to date. They attribute lack of progress to lack of sophistication of the extraction techniques as well as source information that is inconsistent, circular, or missing crucial (often common sense) information.

Other work targets encoding WordNet glosses as axioms in first-order logic (Harabagiu et al. 1999; Rus 2005). The logical information is extracted from syntactically processed glosses using straightforward structural analysis (for example mapping the NP in a subject position to an AGENT role), which limits the depth of the semantic representation. In addition, the representation does not lend itself to using reasoning processes in order to clean up and refine the conceptual definitions.

Our approach addresses many of the shortcomings in previous work. TRIPS gives us a well developed core deep language understanding system that provides a substantial processing boost beyond what is available with other off-the-shelf tools. Furthermore, its ontology defines key concepts and their participant roles in a hierarchy. Each word is associated with a concept and linking templates that align the semantic relations with syntactic structure during parsing. The TRIPS NL framework helps us overcome the bootstrapping problem that limits the depth of information extracted with earlier approaches.

Concluding Remarks

A critical obstacle preventing the constructing of human-level cognitive agents is the lack on common-sense knowledge that it would need to understand language and reason about the world. We hope we have presented a convincing case that extraction of high quality knowledge from sources such as WordNet is a feasible goal to pursue.

While the example we presented in this paper is small, and based on a particularly suitable subset of WordNet, we are currently performing a significantly larger experiment based on approximately 7500 synsets that are likely to be related to cooking. In order to construct an effective knowledge base, we are working on a number of the challenges presented earlier. We plan to evaluate this cooking KB by having subjects pose queries about cooking (based on a provided set of templates) to both the system and to a human using a web-based system, and then scoring the resulting answers for correctness.

References


V. Rus. A Method to Generate Large Common-Sense Knowledge Bases from Online Lexical Resources. FLAIRS 2005.

