An operator-based account of semantic processing

C. Anton Rytting (rytting@ling.ohio-state.edu)
Department of Linguistics
The Ohio State University

Deryle Lonsdale (lonz@byu.edu)
Department of Linguistics
Brigham Young University

Abstract. This paper explores issues of psychological plausibility in modeling natural language understanding within Soar, a symbolic cognitive model. It focuses on constructing syntactic and semantic representations in simulated real time, with particular emphasis on word sense disambiguation (WSD). We discuss (i) what level of WSD should be modeled and (ii) how to use resources such as WordNet to inform these models. A preliminary model of coarse-grained WSD is included to show how syntactic, semantic, and other knowledge sources interact in Soar. Finally, we explore issues of interleaving, learning, and integrating other WSD approaches with Soar's native model of learning.

Keywords: Soar, WordNet, word sense disambiguation, syntax/semantics interface, cognitive modeling

1. NL-Soar, WordNet and WSD

This paper argues that the principled use of WordNet within an operator-based framework yields interesting results in modeling semantic processing in general, and particularly where issues of word sense disambiguation are concerned. It outlines the system's semantic processing module including the use of corpus data in determining semantic constituent compatibility constraints. Current and projected research issues are explored, including further integration with other semantic processing systems and the testing of empirical claims regarding the degree of syntax/semantics interaction. Finally, the role of learning in modeling systems is discussed.

1.1. NL-Soar

Natural-Language Soar (NL-Soar) is a natural-language extension of Soar, a symbolic cognitive modeling system (Newell, 1990). The ultimate purpose of NL-Soar is to perform large-scale, robust language comprehension and production in a psychologically plausible way. Since modeling language use is a complicated problem, the NL-Soar system (as with other Soar systems) uses a goal hierarchy to manage complexity: top-level problems can be decomposed into component subproblems, these can be further split into smaller subgoals, and so on. In this paper the goals discussed will involve the construction of semantic representations for English sentences that enter the system.

One aim of NL-Soar is to model language comprehension as a real-time process. NL-Soar (like humans) receives each input sentence word by word, and each word must be processed (that is, integrated into the larger syntactic and semantic representations) before it decays from a temporary perceptual buffer. NL-Soar must therefore begin syntactic and semantic processing incrementally, without knowing the whole sentence beforehand. Observable effects follow; for example, its syntactic module has been used to predict the point of comprehension failure on garden-path sentences (Lewis, 1993).

NL-Soar is implemented within a typical operator-based framework. This means that operators provide the basic mechanism that drives processing in the system. Operators can be viewed as the means of executing system actions that are initiated when certain preconditions are satisfied. These actions operate on the agent's world model and effect changes to attributes within the model. As sequences of operators are proposed, compared, selected, fired, and terminated one at a time, complex patterns of behavior result. Though there are many types of operators in NL-Soar, this paper focuses on operators that connect together pieces of semantic (or conceptual) structure. In particular, it examines how the operator-based approach provides a flexible framework for modeling and testing various assumptions and theories about human semantic processing.

1.2. WordNet and NL-Soar

Because the NL-Soar system processes natural language sentences, it must have at its disposal information about words. A system lexicon provides knowledge about English words, their parts of speech, and associated morphological, syntactic and semantic information. Until recently, words were added to the NL-Soar English lexicon on a rather ad-hoc and as-needed basis; no systematic method was pursued to assure wide-scale coverage and consistency across entries in the system's lexicon. This was certainly true for semantics: previous versions of NL-Soar used a well-known set of eight primitive semantic categories (Jackendoff, 1990), but had no principled method for assigning new vocabulary into these eight classes.

To provide more principled and thorough coverage for the system's lexicon, NL-Soar has recently been fitted with the WordNet lexical knowledge base (Miller, 1993). This has greatly enhanced NL-Soar's capabilities, but at the same time adds a number of complications of both practical and theoretical import (Lonsdale and Rytting, 2001).

Originally designed and organized according to psycholinguistic principles, WordNet has a number of advantages over traditional lexicons in modeling the human lexicon. Most of its 100,000 or so[1] different concepts (or synonym sets) are arranged into a semantic hierarchy. At the top of WordNet's basic hierarchy are 45 lexical files or "unique beginners", originally formulated for managing the lexicographers' task: 26 noun classes, 15 verb classes, and 4 classes for other parts of speech (Fellbaum, 1990; Miller, 1990; Fellbaum, 1998). Examples of these classes are `v-change`, `v-communication`, `n-artifact`, and `n-food`. These WordNet semantic categories now serve as a basis for NL-Soar's semantic processing.

Since the relation between words and concepts is many-to-many, there must be a way to determine, given a word in its context, which concept is the one being conveyed—in other words, what sense of the word is being used. This task is known as word sense disambiguation (WSD), and constitutes a core area of functionality in NL-Soar semantic processing.

1.3. Word Sense Disambiguation in NL-Soar

WSD is still an active area of research within computational linguistics (Kilgarriff and Palmer, 2000). However, it is not clear to what extent it is really a problem within psycholinguistics, and how much it is a carry-over from traditional lexicography. Some hold that word senses do not really exist as discrete entities, and hence that WSD is a manufactured problem (Melby, 1995; Kilgarriff, 1997; Hanks, 2000).

Psycholinguistic research also provides some cautionary evidence that WSD (and even part-of-speech disambiguation) may not always be a necessary part of human language comprehension (Wilks, 2000). For example, it has been shown that priming across word senses, an experimental effect commonly used to support the existence of word senses in the mind, is not always reliable. The use of words in a less-frequent context may still prime more frequent senses of the word, but more frequent contexts will not always prime less frequent senses (Williams, 1992). This suggests that less-frequent word senses may be stored as extensions of more central definitions, rather than as separate entities.

Another argument against word senses as an inherent aspect of lexical knowledge is the lack of agreement between naïve native speak-

[1] NL-Soar currently uses version 1.6 of WordNet. Subsequently available releases will provide even more types of semantic information to NL-Soar.

ers. On the one hand, word sense tagger reliability with trained experts reaches as high as 95% even for fine-grained senses (Kilgarriff and Palmer, 2000). However, the agreement between naïve speakers is significantly lower, less than 80% (Fellbaum et al., 1998).

These studies suggest that fine-grained WSD may not be an appropriate problem for a psycholinguistic model. Nevertheless, some account of the phenomenon is necessary in order to construct a workable semantic model given a large, polysemous vocabulary. If fine-grained WSD is not feasible or plausible, an approximate model of coarse-grained WSD still seems advisable (Wilks, 1997).

Moreover, some research supports limited, course-grained polysemy in the mental lexicon. The results of one sentence-based semantic classification experiment argue for the existence of a few stable word senses (typically around three) for many nouns, rather than an exhaustive fine-grained inventory (Jorgensen, 1990).

This suggests that it would be better to model disambiguation between a few broadly defined senses of each word than to use dictionary definitions; or alternatively, to disambiguate instances of high-level ontological concepts. This latter approach might be termed semantic class disambiguation (SCD), and is exemplified by the use of Roget's categories (Yarowsky, 1992), and the use (Krymolowski and Roth, 1998) of CoreLex, a set of 126 semantic classes derived from WordNet (Buitelaar, 1998).

2. Semantic representation in NL-Soar

The current version of NL-Soar follows this second strategy, using WordNet's 45 unique beginners as its semantic classes. This first approximation at a semantic model enables exploiting WordNet's vocabulary coverage without committing to the plausibility of its fine-grained concepts. As the WSD modeling granularity issue evolves, the general techniques tested here will be refined.

The semantic structure that is created when an utterance is run in the NL-Soar system is called a Lexical-Conceptual Structure (LCS). An LCS illustrates relationships that exist between the argument(s) and predicate of an utterance. The usual components of an NL-Soar LCS include the predicate of the utterance (the verb), and the argument(s) that the predicate takes. The arguments are of two types: external and internal. The external argument generally refers to the subject of the sentence and it is called external because it is outside the domain of the verb phrase. On the other hand, an internal argument is one that occurs within the domain of the verb phrase, usually an object. Besides arguments, properties may also be linked with an object and usually represent adjectives, adverbs, and prepositional-phrase modifiers.

A semantic link in NL-Soar consists of three parts: an assigning entity, a receiving entity, and a semantic role or link type (for example external, internal, or property). Consider NL-Soar's semantic model of the sentence *The dog barked.*, which could be schematically represented thus:

```
bark.v-body (assigner)--External--> dog.n-animal (receiver)
```

NL-Soar's semantic module currently uses three criteria to link concepts: syntactic and morphological constraints, frequency rankings, and semantic pairing constraints.

First, NL-Soar processes the syntax of an utterance and then constrains the possible semantic readings to those consistent with the syntactic model. Proposed links reflecting relevant criteria are preferred over those which do not. Syntax thus plays a role in SCD (Lonsdale and Rytting, 2001); section 5 below discusses further implications.

Research indicates that word sense frequency may provide a plausible default ranking strategy (Fellbaum et al., 1998). When given a frequency-ordered list, subjects often pick the first sense which reasonably fits the context. This suggests that in absence of other factors (such as discourse context) word sense frequency is an important criterion for assigning meaning to a word.

For a given word, NL-Soar ranks the frequency of each semantic class according to the ranking of most frequent WordNet sense in that class. For example, the word *tooth* has 5 WordNet senses: 1 and 3 fall under the semantic class `n-body`; senses 2 and 5 fall under `n-artifact`; and sense 4 under `n-act`. Hence, NL-Soar's semantic module assigns *tooth* to these four semantic classes, with `n-body` treated as most frequent, followed by `n-artifact` and `n-act`.

The base frequency of semantic classes is not sufficient to determine the most plausible semantic links, however. In the absence of a robust model of real-world knowledge, NL-Soar uses corpus-derived "semantic collocation constraints" determined by frequencies of pairings of certain semantic classes. The strategy is similar to that of (Resnik, 1997; Ciaramita and Johnson, 2000); however, the frequencies are (currently) calculated over relations rather than over specific nouns and verbs (Agirre and Martinez, 2001b).

Canonical semantic collocations were found for NL-Soar by searching SEMCOR, a WordNet-sense-tagged version of the Brown Corpus (Francis and Kučera, 1982), for examples of each verb class. For each verb class, a sample of approximately 50 verb instances were gathered. For each of these instances, the semantic classes for the nouns that fill the verb's external and internal roles were recorded. For example, the following semantic pairings were found to be the most common for the semantic classes `v-body`:

```
External v-body: n-artifact, n-body, n-person
Internal v-body: n-artifact, n-attribute, n-body
```

3. Semantic Processing in NL-Soar

As a given word undergoes semantic processing, all of its possible semantic classes and parts of speech are considered, in order of frequency, as determined above. The first one to pass the selectional pairing constraints is accepted as the best choice and fitted into the semantic model.

However, NL-Soar does not perform SCD on individual words directly. Since the semantic constraints license concept pairings between noun classes and verb classes, NL-Soar must disambiguate noun-verb pairs as a unit, not separately.

For example, consider the sentence *My tooth hurts*. The appropriate (and most frequent) semantic pairing between *tooth* and *hurt* is:

```
(1) hurt.v-body --External-->  tooth.n-body.
```

Operators also propose other possible links:

```
(1') hurt.v-body --Internal-->  tooth.n-body.
(2) hurt.v-body --External-->  tooth.n-artifact
(3) hurt.v-change --External-->  tooth.n-body
(4) hurt.v-change --External-->  tooth.n-artifact (etc.)
```

but these are not considered first, since they either do not match the syntax (in case of 1') or are not most frequent. Since `(v-body, External, n-body)` is a relatively common semantic pairing, it is licensed by the constraints, and the pairing (1) is linked with the ongoing semantic model.

Although the various potential semantic links are initially considered in parallel, semantic constraints are tested serially in NL-Soar. Word sense frequency determines which links are tested first. If both words in the link are semantically ambiguous, it is unclear which word's semantic-class frequency ranking should carry more weight. A number of criteria could be used: e.g., part of speech, syntactic role (assigner vs. receiver), and degree of polysemy. This study assumes temporal ordering as the key factor. Since the earlier of the two words being linked is (in the absence of extra-sentential context) less constrained by previously heard context than the later word, its base-line word sense frequency ranking should apply more pertinently. Hence, the most frequent sense of the first ambiguous word should most strongly constrain the acceptable word senses of later words.

This claim is not essential to the model but rather another working hypothesis in need of empirical verification. There seems to be preliminary neurolinguistic support: an event-related potential (ERP) study shows that in normal, congruous sentences, N400 anomaly effects for low-frequency words are greatest sentence-initially, and decrease thereafter (Van Petten, 1995). See section 6 for further implications.

4. Specifying and employing constraints

The approach described above has been implemented in NL-Soar and is undergoing continuous refinement. As semantic collocation information is found from SEMCOR, it is encoded into the system as constraints that permit (or alternatively disallow) semantic attachments. This section explores examples where SEMCOR constraints have been implemented and explains how the information is used. For ease of discussion and in order to reflect current system capabilities, a number of syntactic simplifications have been made to these sentences, in order to focus on lexical and semantic issues. All simplified sentences consisted of noun phrases and finite verbs. These results must be interpreted cautiously, since the semantic constraints are not complete for all verb classes. However, NL-Soar provides an acceptable parse on many types of sentences, performing much better than chance when the number of variables involved are considered.

In order to first understand how operators assure construction of semantic representations, three sentences will be contrasted. The first, *The woman yawned.*, has two senses (`n-person` and `n-group`) for the noun and two (`v-stative` and `v-body`) for the verb. In this case the system, having already processed the syntax, tries to link together the most frequent senses of each word. Figure 1 shows how this is done, showing relevant semantic operators (with a running total of operator firings, or cycles, indicated by integers on the left). First, the system tries to fuse (or connect together) the concepts `yawned.v-body` and `woman.n-person`, with the latter in the External slot of the former. Several constraints must license this attachment, though, so various operators (cycles 149 through 161) check these constraints. Since all constraints (e.g. word order, semantic category, uniqueness of role assignment, and consistency with syntactic facts) are satisfied with this

```
147:   O: O136 try(fuse(yawned.v-body--Ext-woman.n-person)
148:   ==>S: S130 (operator no-change)
149:      O: C264 (check sense-check)
150:      O: C278 (check order-constraint)
151:      O: C276 (check semcat-check)
152:      O: C274 (check semcat-check)
153:      O: C272 (check semcat-check)
154:      O: C270 (check semcat-check)
155:      O: C268 (check semcat-check)
156:      O: C266 (check duplicate-fusion)
157:      O: C262 (check category-check)
158:      O: C258 (check check-syntax)
159:      O: C256 (check adjacency)
160:      O: C260 (check not-duplicate-relation)
161:      O: O139 (constraint-success)
162: O: C253 fuse(yawned.v-body--Ext-woman.n-person)
```

Figure 1. Linking together the semantics for the sentence *The woman yawned.*

proposed course of action, the system executes the fuse operator in the last step, thus attaching the two concepts together.

The relevant portion of processing for the second sentence, *The canyon yawned before us.*, is shown in Figure 2. Note that here, in operator cycles 141-145, the system first tried to link `n-object` in the external slot of the `v-body` concept. This violates a semantic constraint (i.e. physical objects cannot perform body verbs) in cycle 145, so the system aborts pursuing this possibility in favor of the next hypothesis: the `n-object` links in the `v-stative` sense's external slot. After extensive constraint checking, the attachment is licensed and performed in cycle 156.

Now consider the third sentence, *The chair yawned.*, where the verb has the same senses as before, but *chair* has three senses: `n-act`, `n-person` (a short form for *chairman* or *chairwoman* or *chairperson*), and `n-artifact`. Here (see Figure 3) the artifact reading and the `n-act` reading (the two most common senses for *chair*) are tried first with the body verb (the most common sense for *yawned*), incompatible choices as detected in cycles 178 and 183. Finally, the system tries the least common noun sense (`n-person`) with the verb. This attachment clears several constraints and gets constructed in cycle 194.

Brief mention of a few examples is instructive in appreciating the issues that arise from using WordNet and in particular SEMCOR as a data source for NL-Soar. Semantic constraints were tested for external

```
141:   O: O124 try(fuse(yawned.v-body--Ext-canyon.n-object)
142:   ==>S: S125 (operator no-change)
143:      O: C252 (check sense-check)
144:      O: C266 (check order-constraint)
145:      O: C264 (check semcat-check)
146:   O: O123 try(fuse(yawned.v-stative--Ext-canyon.n-object)
147:   ==>S: S126 (operator no-change)
148:      O: C279 (check sense-check)
149:      O: C283 (check order-constraint)
150:      O: C281 (check duplicate-fusion)
151:      O: C277 (check category-check)
152:      O: C273 (check check-syntax)
153:      O: C271 (check adjacency)
154:      O: C275 (check not-duplicate-relation)
155:      O: O126 (constraint-success)
156: O: C242 fuse(yawned.v-stative--Ext-canyon.n-object)
```

Figure 2. Linking together the semantics for the sentence *The canyon yawned before us.*

```
174:   O: O208 try(fuse(yawned.v-body--Ext-chair.n-artifact)
175:   ==>S: S146 (operator no-change)
176:      O: C364 (check sense-check)
177:      O: C378 (check dummy-constraint)
178:      O: C376 (check semcat-check)
179:   O: O210 try(fuse(yawned.v-body--Ext-chair.n-act)
180:   ==>S: S147 (operator no-change)
181:      O: C391 (check sense-check)
182:      O: C405 (check dummy-constraint)
183:      O: C403 (check semcat-check)
184:   O: O207 try(fuse(yawned.v-body--Ext-chair.n-person)
185:   ==>S: S148 (operator no-change)
186:      O: C418 (check sense-check)
187:      O: C422 (check dummy-constraint)
188:      O: C420 (check duplicate-fusion)
189:      O: C416 (check category-check)
190:      O: C412 (check check-syntax)
191:      O: C410 (check adjacency)
192:      O: C414 (check not-duplicate-relation)
193:      O: O221 (constraint-success)
194: O: C344 fuse(yawned.v-body--Ext-chair.n-person)
```

Figure 3. Linking together the semantics for the sentence *The chair yawned.*

roles; internal roles have not yet been addressed (though work on this topic will follow). Each sentence of interest was taken from SEMCOR, and most were simplified to abstract away from nonessential material without changing the semantic classes involved. For example, a SEMCOR sentence such as *The antagonists were instructed to behave in such ways as to upset the manager and get him to operate in a manner for which he had been previously criticized.* was simplified to *The antagonists behave annoyingly.*. For each sentence, the SEMCOR tags were compared with those assigned by NL-Soar. Additional thematic role information which is not provided in SEMCOR (i.e. E(xternal) for agent, or deep subject, and I(nternal) for patient/theme, or deep object), but which is calculated by NL-Soar, is also indicated in the examples that follow. An asterisk indicates that the role is left unfilled (e.g. an optionally empty Internal role). Mismatches are indicated with italics. Linkages were categorized into one of four classes: (1) match with SEMCOR, (2) acceptable mismatch (i.e., a plausible reading), (3) unacceptable mismatch, and (4) fail, where the system provides no semantic interpretation.

Consider, for example, the (simplified) sentence *The exercise exhausts the child.*. NL-Soar's semantic linkage exactly reflects that of SEMCOR's:

`exercise.n-act-Extern-exhaust.v-body-Intern-child.n-person`

The class `exhaust.v-body` is the most frequent class for *exhaust* (and also the most frequent sense, out of 5 senses). Constraints that follow from semantic processing and semantic cooccurrence licensing allow both External and Internal matching[2].

Another sentence, *Clergy wear grey.*, illustrated another interesting challenge—the familiar training/testing problem. When extracting information from processing a portion of SEMCOR, one particular collocation was not encountered: the class `n-group` never showed up in an External role for `v-body`. Yet when testing NL-Soar on sentences from elsewhere in SEMCOR, three such sentences were encountered. This type of problem is easily overcome by fiat; the addition of one more precondition to an existing rule to relax this constraint improves results tremendously. However, a more satisfying solution would be to

[2] An interesting future extension would be to assign the External and Internal with STIMULUS and EXPERIENCER theta-roles, or otherwise relating the External `n-act` with the Internal `n-person`, since surely their co-occurrence is not independent. One weakness of the current sampling mechanism is to treat External and Internal roles independently, and hence lose this potentially useful information. On the other hand, trying to build in too much information could lead to a sparse data problem, unless more automatic methods of sampling are developed.

discover what types of general underspecification schemata, combined with what types of (semi-labelled) input, would allow NL-Soar's built-in learning mechanism to learn the connection between `n-group` and `n-person` on its own. (See section 7 for more details on learning.)

One example of an unacceptable mismatch may be due to the quality of SEMCOR tagging instances. The sentence *Antagonists behave badly.* has the following SEMCOR and NL-Soar linkages:

```
SEMCOR: antagonist.n-group-Ext- behave.v-body-Int-*
NL-Soar: antagonist.n-body-Ext- behave.v-body-Int-*
```

Here NL-Soar has taken the `v-body` class for *behave*, and consequently prefers an `n-body` linkage for the subject *antagonist*; even though the `n-group` class is more frequent, collocation constraints do not prefer linking `n-group` with `v-body`. On the other hand, *behave* has another class, `v-social`; if SEMCOR would have had this arguably more appropriate sense tagged instead, NL-Soar would have been able to perform the proper linkage with `n-group`.

WordNet's exhaustiveness often causes homonymy-related problems; words such as *at*, *a*, *an*, *are*, etc. have nominal readings that are not readily identifiable by non-lexicographers. Most of these lexical, open-class readings (e.g. *are* as a unit of measure) complicate syntactic and semantic processing, and consequences were seen while testing SEMCOR-derived sentences. The solution is straightforward in NL-Soar: a rule was introduced that automatically prefers the functional reading for obscure homographs like these.

Finally, consider the sentence *The epidemic infected the citizens.*, an instance where the frequency as a measure of the primary or "central" word sense fails. Note the SEMCOR and NL-Soar linkages:

```
SEMCOR: epidemic.n-event-E-infect.v-body-I-citizen.n-person
NLSoar: epidemic.n-event-E-infect.v-cognition-I-citizen.n-person
```

Intuitively, the "central" meaning of infect is the medical (or `v-body`) reading glossed as: "communicate a disease to". However, a metaphorical `v-cognition` reading "fill with revolutionary ideas" is listed as the most frequent one in WordNet. Oddly enough, the original context for the test sentence (*"It is like a mysterious epidemic, which ... spreads ... through the whole town until all have been infected by it."*) was also embedded in a simile which, in the text's wider context, referred to the spreading of (dangerous) ideas. While it is unsurprising that the detection and processing of metaphor and figurative language presents difficulties, it is striking that WordNet's frequency rating points out the correct direction of the text's larger simile, even though it mispredicted the literal annotation found in SEMCOR.

5. Modeling: what is the target?

The approach sketched above is, of course, not the only one possible. Instead, it was adopted as a baseline strategy to serve while the basic semantic processing architecture set in place. In fact, several aspects of psycholinguistic groundwork still need to be done in the areas of WSD and on-line semantic processing; comprehensive modeling must either await such results, or else propose models that might inspire (or challenge) experimentalists to verify or refute claims put forward by speculative modeling efforts.

As computational and architectural issues remain an area of focus, an operator-based approach can support whatever solution might be followed in addressing these issues. For example, a determination of the most appropriate set of off-the-shelf semantic classes (or indeed entire ontologies) would be useful; possibilities include the Roget and CoreLex categories mentioned above. Several off-the-shelf WSD algorithms also remain to be explored; a partial review of the types of information relevant to WSD is found in (Agirre and Martinez, 2001a).

As the situation stands now, "training" and "performance" are artificially separated as hand-sampling of collocation patterns replaces (or rather, holds a place for) the processing mechanisms native to Soar. A better way to look at the problem of Soar processing is the combination of intelligent use of heuristics and remembering the results from applying these heuristics. The Soar framework is specifically designed to support a panoply of strategies and processing modalities, and to manage the decision process involving which to use when.

For example, one (or even several) of the previously mentioned WSD techniques could be accessed as strategy when NL-Soar is unsure of the most appropriate word in context. NL-Soar could first search for "give-away" bigrams (as in (Pedersen, 2001)). If no such bigrams are in the previously heard context, then other techniques such as those mentioned above could be used. If insufficient data is available for a particular word, then a strategy such as (Clark and Weir, 2001) could be used to find a more suitable level on the hierarchy on which to repeat the reasoning process. There need be no assumption of a single WSD technique appropriate for all occasions.

The relationship between syntactic and semantic processing has been a subject of considerable debate within psycholinguistics. Two extreme positions exist; the first assumes complete precedence of syntax over semantics. Doubt has been cast on the plausiblity of this extreme; consider this popular example: *After finding the book on the atom, Kim decided that the library really wasn't as bad as people had been claiming.* (Hirst, 1987)

It seems clear that *book* and *atom* must be matched with real-world concepts (along with the most plausible meaning of *on*) before the complete sentence is heard. While no-one likely takes the separation to this extreme, it would appear that some models might adopt it for reasons of simplicity.

The other extreme is that of complete two-way interaction between syntax and semantics. A considerable degree of interaction is claimed by the constraint satisfaction approach (MacDonald et al., 1994; Trueswell et al., 1994; Burgess and Lund, 1997) and assumed in some connectionist approaches. However, this claim is still controversial, and other works provide evidence against it (Frazier and Fodor, 1978; Ferreira and Clifton, 1986; Clifton et al., 1991). While much psycholinguistic research remains to be made, the parallel development of flexible and versatile frameworks will help current and future modeling efforts.

6. Interleaving and Interactivity

NL-Soar's use of operators, which can be interleaved, allows it to model an intermediate position between these two extremes. While NL-Soar currently assumes that syntax guides semantics, this need not be a temporal ordering. Rather, through interleaving of the syntactic and semantic processes, the semantics module may begin processing the syntactic module's early results while the syntactic module processes later parts of the sentence. Misanalyses in syntax and semantics can be corrected through "snip" and "semsnip" operators, respectively.

This interaction of syntax and semantics is easily seen during processing of the sentences *Zebras have hooves.* and *Zebras have sneezed.*. When the second word enters the system *Zebras have...*, there is momentary ambiguity[3] concerning the function and role of the word *have*; the system perfers the `v-possession` reading. Therefore, the sentence *Zebras have hooves.* proceeds according to principles described above. In this case, as shown in the trace in Figure 4, the external argument is correctly attached after constraints are satisfied (cycles 116-126), and then the next word enters the system and is accessed (cycle 133). After the word is incorporated into the syntactic tree, the semantic attachment is also made (cycles 178-189).

Consider, though, the other sentence: *Zebras have sneezed.* where the possessive reading is not correct for *have*, which serves here as

[3] In spoken language understanding, this ambiguity may be bypassed by listeners' attention to the duration or stress on "have"—usually noticeably shorter for the auxiliary form. On the other hand, spoken language introduces other ambiguities not found in written text.

```
116:   O: O175 try(fuse(have.v-possession--Ext-zebras.n-animal)
117:   ==>S: S120 (operator no-change)
118:      O: C276 (check sense-check)
119:      O: C280 (check dummy-constraint)
120:      O: C278 (check duplicate-fusion)
121:      O: C274 (check category-check)
122:      O: C270 (check check-syntax)
123:      O: C268 (check adjacency)
124:      O: C272 (check not-duplicate-relation)
125:      O: O177 (constraint-success)
126: O: C266 fuse(have.v-possession--Ext-zebras.n-animal)

133:  O: A171 (access word: 'hooves' spkr: user)

178:   O: O248 try(fuse(have.v-possession--Int-hooves.n-animal)
179:   ==>S: S169 (operator no-change)
180:      O: C420 (check sense-check)
181:      O: C426 (check dummy-constraint)
182:      O: C424 (check semreceiver-follows)
183:      O: C422 (check duplicate-fusion)
184:      O: C418 (check category-check)
185:      O: C414 (check check-syntax)
186:      O: C412 (check adjacency)
187:      O: C416 (check not-duplicate-relation)
188:      O: O249 (constraint-success)
189: O: C411 fuse(have.v-possession--Int-hooves.n-animal)
```

Figure 4. Linking together the semantics for the sentence *Zebras have hooves.*

an auxiliary. Figure 5 shows that, as with the other sentence, cycles 116-126 establish the possessive reading in the semantics. But now when the word *sneezed* is accessed (cycle 171), the syntax is forced into a reanalysis since the possessive construal is no longer tenable. A syntactic snip operator (cycle 157) removes the NP "zebras" from the (possessive) verb and re-links it to the auxiliary verb (cycles 164-165). Crucially, the semantic representation must also be snipped in tandem; this is performed by the semsnip operator in cycle 177. This semsnip removes the erstwhile External argument of the `v-possession` sense of *have*, thus making the subject *zebras* available for another semantic attachment. Accordingly, in cycles 182-196 the system attempts to attach `zebras.n-animal` to the External argument slot of `sneezed.v-body`; after several successful constraint checks, the attachment is made.

```
116:   O: O175 try(fuse(have.v-possession--external-->zebras.n-animal)
117:   ==>S: S120 (operator no-change)
118:      O: C276 (check sense-check)
119:      O: C280 (check dummy-constraint)
120:      O: C278 (check duplicate-fusion)
121:      O: C274 (check category-check)
122:      O: C270 (check check-syntax)
123:      O: C268 (check adjacency)
124:      O: C272 (check not-duplicate-relation)
125:      O: O177 (constraint-success)
126: O: C266 fuse(have.v-possession--external-->zebras.n-animal)

133: O: A171 (access word: 'sneezed' spkr: user)

157: O: O209 snip(zebras)
164:   O: O213 try(link(have.i--spec-->zebras.n)
165: O: C365 link(have.i--spec-->zebras.n)

177: O: O229 (semsnip)

182:   O: O295 try(fuse(sneezed.v-body--external-->zebras.n-animal)
183:   ==>S: S177 (operator no-change)
184:      O: C480 (check sense-check)
185:      O: C494 (check dummy-constraint)
186:      O: C492 (check semcat-check)
187:      O: C490 (check semcat-check)
188:      O: C488 (check semcat-check)
189:      O: C486 (check semcat-check)
190:      O: C484 (check semcat-check)
191:      O: C482 (check duplicate-fusion)
192:      O: C478 (check category-check)
193:      O: C474 (check check-syntax)
194:      O: C472 (check adjacency)
195:      O: C476 (check not-duplicate-relation)
196:      O: O297 (constraint-success)
197: O: C470 fuse(sneezed.v-body--external-->zebras.n-animal)
```

Figure 5. Linking together the semantics for the sentence *Zebras have sneezed.*

Snips (and the associated semsnips) can only be employed to a certain point, however; beyond that, sentences are predicted to result in garden pathing (Lewis, 1993). NL-Soar's framework is thus able to account for a great deal of interaction between modules while preserving the separation commonly assumed; however, it could be modified to accommodate full interaction between modules, should further evidence come out in favor of that hypothesis.

This leads into a further claim, mentioned in section 3, that frequency effects play a larger role in the disambiguation of early words than later words in an utterance (outside of prior disambiguating context). This claim seems consistent with either the strong interactive (constraint satisfaction) stance or with NL-Soar's current interleaving framework. It predicts the existence of "semantic garden paths" depending on early rejection of a low-frequency meaning of the first word, followed by late disambiguating material. Again, psycholinguistic exploration of this area would be valuable to establish a target for semantic modeling efforts.

It is expected that, once other issues regarding other aspects of WSD are settled, NL-Soar can be a useful tool in investigating this claim, in conjunction with corresponding psycholinguistic experiments.

7. Transfer and learning

One important property of operators is that their effects can be cast in terms of hierarchical subgoal satisfaction. When the knowledge about how to attain a goal (such as computing a semantic linkage) is lacking in a holistic sense, NL-Soar can subgoal and test various possibilities in an attempt to achieve the top-level goal. This may involve extensive levels of decomposition, each of which may in turn involve lower-level operators, constraint checking, and deliberation. Though this process may be costly in terms of processing time and memory requirements the first time it is carried out, the system also provides a significant capability for such situations—machine learning. Results (whether successful or not) can be "chunked up" automatically during subgoaling, and the results will then be available for subsequent processing. This enables the so-called recognitional mode of operation in certain contexts.

For example, consider the sentence *Zebras have hooves.* which was used earlier. When NL-Soar encounters this sentence without any prior-learned language ability, its processing requires 135 operator cycles, 3.5 seconds of CPU time, 8272 rule firings, just over 18,000 working memory element changes, and a mean working memory size of 1670 elements. On the other hand, once this sentence (or some similar one) has been processed beforehand and the learned results are available to the system in the form of chunked operators, significantly less processing resources are required. Specifically, processing the sentence recognitionally demands only 1.5 seconds of CPU time, a mere 11 operator cycles, 1797 rule firings, 2647 working memory changes, and a mean working memory size of 857 elements.

When learned operators for semantic processing are available, they thus leverage experience gained in accepting or rejecting certain pairings of semantic substructure as described in section 2. WordNet semantic classes, along with decisions about their collocational and selectional implementations, can thus be incorporated into the learning process.

Interesting issues which cannot be addressed here involve the granularity of the semantic operators: to what degree should the semantic class categories which are used as linking constraints be generalized? To what extent can the hand-derived corpora be abstracted away from, in order to increase the coverage? Language acquisition research in such areas as categorization and semantic overgeneralization could be relevant in exploring these issues.

An intriguing learning-based direction involving the semantic processing mentioned in this paper will be possible when more than one WSD method is implemented, as discussed earlier. These strategies could even inform one another through memorized chunks and metalearning strategies. For example, if one WSD mechanism is successful at identifying a unique word sense at a high degree of confidence, then it could train other strategies, perhaps by remembering salient bigrams for future expansion of Pedersen's bigram heuristic. The system may even be able to learn when some heuristics (e.g., one-sense-per-discourse (Gale et al., 1992) or one-sense-per-collocation (Yarowsky, 1993)) are valid, when they should be avoided, and when one strategy is preferable to another. To date none of these strategies has yet been implemented in NL-Soar, but each should be feasible within the framework.

Another advantage NL-Soar derives from the operator-based architecture is the possibility of transfer. Tasks which leverage similar low-level processes can share operators for these processes; hence, crossmodal application of operators may account for facilitation and inhibition across tasks, and allow the generalization of semantic or syntactic facts beyond the specific context of a learning corpus. Bootstrapping of semantic knowledge through supervised learning is thus one avenue of future work.

8. Conclusions

This paper has surveyed the semantic processing component of NL-Soar, a computational model of human language processing based on an operator framework. Operators provide the flexibility to model language use according to various theoretical and empirical approaches. The system currently performs semantic model construction based on fairly commonly held assumptions. The semantic processing module containts a model of coarse-grained word sense disambiguation, currently based on WordNet's top-level semantic classes and guided by cooccurrence constraints observed in hand-extracted frames from SEMCOR, a semantically annotated corpus. This process can ultimately be automated via a wide range of possible approaches. Interesting related topics such as learning, interactivity, operator granularity, task-related transfer, and priming are currently being explored.

References

Agirre, E. and D. Martinez: 2001a, 'Knowledge Sources for Word Sense Disambiguation'. In: V. Matousek, P. Mautner, R. Moucek, and K. Tauser (eds.): *Text, Speech and Dialogue, 4th International Conference, TSD 2001, Zelezna Ruda, Czech Republic, September 11-13, 2001, Proceedings*, Vol. 2166 of *Lecture Notes in Computer Science*. Springer.

Agirre, E. and D. Martinez: 2001b, 'Learning class-to-class selectional preferences'. In: W. Daelemans and R. Zajac (eds.): *Proceedings of CoNLL-2001*. pp. 15–22, Toulouse, France.

Buitelaar, P.: 1998, 'CORELEX: Systematic Polysemy and Underspecification'. Ph.D. thesis, Brandeis University, Computer Science Department.

Burgess, C. and K. Lund: 1997, 'Modelling Parsing Constraints with High-dimensional Context Space'. *Language and Cognitive Processes* **12**(2/3).

Ciaramita, M. and M. Johnson: 2000, 'Explaining away ambiguity: Learning verb selectional preferences with Bayesian Networks'. In: *Proceedings of the 18th International Conference on Computational Linguistics*. pp. 187–193.

Clark, S. and D. Weir: 2001, 'Class-based probability estimation using a semantic hierarchy'. In: *WordNet and other lexical resources: applications, extensions, and customizations*. Association for Computational Linguistics. NAACL-2001.

Clifton, Jr., C., S. Speer, and S. P. Abney: 1991, 'Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions'. *Journal of Memory and Language* **30**, 251–271.

Fellbaum, C.: 1990, 'English verbs as a semantic net'. *International Journal of Lexicography* **3**(4), 278–301.

Fellbaum, C.: 1998, *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fellbaum, C., J. Grabowski, and S. Landes: 1998, 'Performance and confidence in a semantic annotation task'. In: C. Fellbaum (ed.): *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press, pp. 217–238.

Ferreira, F. and C. J. Clifton: 1986, 'The independence of syntactic processing'. *Journal of Memory and Language* **25**, 348–368.

Francis, W. N. and H. Kučera: 1982, *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Frazier, L. and J. D. Fodor: 1978, 'The sausage machine: A new two-stage parsing model'. *Cognition* **6**, 291–325.

Gale, W., K. Church, and D. Yarowsky: 1992, 'One sense per discourse'. In: *Proceedings of the DARPA Speech and Natural Language Workshop*. New York, pp. 233–237, Harriman.

Hanks, P.: 2000, 'Do word meanings exist?'. *Computers and the Humanities* **34**(1-2), 205–215.

Hirst, G.: 1987, *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: CUP.

Jackendoff, R.: 1990, *Semantic Structures*, Current Studies in Linguistics Series. MIT Press.

Jorgensen, J. C.: 1990, 'The Psychological Reality of Word Senses'. *Journal of Psycholinguistic Research* **19**(3), 167–190.

Kilgarriff, A.: 1997, '"I don't believe in word senses"'. *Computers and the Humantities* **31**, 91–113.

Kilgarriff, A. and M. Palmer: 2000, 'Introduction to the special issue on SENSEVAL'. *Computers and the Humanities* **34**(1-2), 1–13.

Krymolowski, Y. and D. Roth: 1998, 'Incorporating Knowledge in Natural Language Learning: A case study'. In: S. Harabgiu (ed.): *Usage of WordNet in natural language processing systems: Proceedings of the workshop*. pp. 121–127. COLING-ACL-98.

Lewis, R.: 1993, 'An Architecturally-based Theory of Human Sentence Comprehension'. Technical Report CMU-CS-93-226, CMU.

Lonsdale, D. and C. A. Rytting: 2001, 'Integrating WordNet in NL-Soar'. In: *WordNet and other lexical resources: applications, extensions, and customizations*. pp. 162–164, Association for Computational Linguistics. NAACL-2001.

MacDonald, M. C., N. J. Pearlmutter, and M. S. Seidenberg: 1994, 'The lexical nature of syntactic ambiguity resolution'. *Psychological Review* **101**, 676–703.

Melby, A.: 1995, *The Possibility of Language*. Amsterdam: John Benjamins.

Miller, G.: 1990, 'Nouns in WordNet: a lexical inheritance system'. *International Journal of Lexicography* **3**(4), 245–264.

Miller, G.: 1993, 'WordNet: A lexical database for English'. In: *Communications of the ACM*, Vol. 38.

Newell, A.: 1990, *Unified Theories of Cognition*. Harvard University Press.

Pedersen, T.: 2001, 'A decision tree of bigrams is an accurate predictor of word sense'. In: *2nd Meeting of the North American Chapter of the Association for Computational Linguistics: Proceedings of the Conference*. pp. 79–86, Association for Computational Linguistics.

Resnik, P.: 1997, 'Selectional Preference and sense disambiguation'. *Proceedings of the ANLP-97 Workshop: Tagging Text with Lexical Semantics: Why, What, and How?*

Trueswell, J. C., M. K. Tanenhaus, and S. M. Garnsey: 1994, 'Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution'. *Journal of Memory and Language* **33**, 285–318.

Van Petten, C.: 1995, 'Words and sentences: Event-related brain potential measures'. *Psychophysiology* **32**, 511–525.

Wilks, Y.: 1997, 'Senses and texts'. *Computers and the Humanities* **31**, 77–90.

Wilks, Y.: 2000, 'Is word sense disambiguation just one more NLP task?'. *Computers and the Humanities* **34**, 235–243.

Williams, J. N.: 1992, 'Processing Polysemous Words in Context: Evidence for interrelated meanings.'. *Journal of Psycholinguistic Research* **21**, 193–218.

Yarowsky, D.: 1992, 'Word-sense disambiguation using statistical models of Roget's categories trained on large corpora'. In: *Proceedings of COLING-92*. Nantes, pp. 454–460.

Yarowsky, D.: 1993, 'One sense per collocation'. In: *Proceedings of the ARPA Human Language Technology Workshop*. pp. 266–271, Morgan Kaufman.