A COMPUTATIONAL THEORY OF VOCABULARY ACQUISITION

William J. Rapaport and Karen Ehrlich

Department of Computer Science and Center for Cognitive Science
State University of New York at Buffalo, Buffalo, NY 14260-2000
{rapaport|ehrlich}@cs.buffalo.edu
http://www.cs.buffalo.edu/sneps/

14 April 1998

Abstract

As part of an interdisciplinary project to develop a computational cognitive model of a reader of narrative text, we are developing a computational theory of how natural-language-understanding systems can automatically acquire new vocabulary by determining from context the meaning of words that are unknown, misunderstood, or used in a new sense. 'Context' includes surrounding text, grammatical information, and background knowledge, but no external sources. Our thesis is that the meaning of such a word *can* be determined from context, can be *revised* upon further encounters with the word, "*converges*" to a dictionary-like definition if enough context has been provided and there have been enough exposures to the word, and eventually "*settles down*" to a "steady state" that is always subject to revision upon further encounters with the word. The system is being implemented in the SNePS knowledge-representation and reasoning system.

This essay is forthcoming as a chapter in Iwańska, Łucja, & Shapiro, Stuart C. (1999), *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language* (Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press). The present version is *Technical Report* 98-05 (Buffalo: SUNY Buffalo Department of Computer Science) and *Technical Report* 98-1 (Buffalo: SUNY Buffalo Center for Cognitive Science).

1 Introduction

As part of an interdisciplinary project to develop a computational cognitive model of a reader of narrative text (Duchan et al. 1995), we are developing a computational theory of how a natural-language-understanding system (either human or artificial) can automatically acquire new vocabulary by determining from context the meaning of words that are unknown, misunderstood, or used in a new sense (Ehrlich 1995),[1] where 'context' includes surrounding text, grammatical information, and background knowledge, but no access to external sources of information (such as a dictionary or a human).

In principle, there are only a handful of ways such a system could learn the meaning of a new word (or revise its understanding of a word): (1) It could look the word up (in, e.g., in a dictionary; cf. Zadrozny & Jensen 1991), (2) it could ask another system what the word means (cf., e.g., Haas & Hendrix 1983, Zernik & Dyer 1987), (3) it could (either by itself or with the help of another system) figure out the meaning by finding a synonym for the unknown word or by locating the meaning of the word in some taxonony or other schema (cf., e.g., Hastings 1994), or (4) it could figure out the meaning entirely from the context in which the word was encountered, with no outside help and no pre-established schema within which to fit it. The last of these is what our system does.

As part of our goal of modeling a reader, it is important to model the ability to *learn* from reading, in particular, to expand one's vocabulary in a natural way while reading, without having to stop to ask someone or to consult a

[1] An earlier version of the present chapter appeared as Ehrlich & Rapaport 1997.

dictionary. A complete lexicon cannot be manually encoded, nor could it contain new words or new meanings (Zernik & Dyer 1987). Text-understanding, message-processing, and information-extraction systems need to be robust in the presence of unknown expressions, especially systems using unconstrained input text and operating independently of human intervention, such as "intelligent agents". For example, a system designed to locate "interesting" news items from an online information server should not be limited to keyword searches—if the user is interested in news items about a certain kind of entity, and the filter detects items about "brachets" (a term not in its lexicon), it should deliver those items as soon as it figures out that a brachet is an entity of that kind. (After reading §2, you will know what kind of entity this is.)

The definition that our system constructs is treated as a current hypothesis about the word's meaning, which can be revised upon successive encounters with the word. Using a formal notion of what a "good" definition is (see §6), our system finds fillers for slots in a definitional frame.

2 An Example

Before describing our system in more detail, it will be useful to see an example of how it works. If you do not know what 'brachet' means, read on; otherwise, try the following experiment with someone who doesn't know what it means.

Suppose that you are reading the 15th-century *Morte Darthur* (Malory 1470) and come upon the following passage (our boldface):

(1) … there came a white hart running into the hall with a white **brachet** next to him, and thirty couples of black hounds came running after them …. (p. 66.)

Suppose that you have no access to a dictionary containing 'brachet' and no one is around whom you can ask what 'brachet' means. Can you figure it out from the context plus your current background knowledge? One subject whom we tried this with guessed

> It's an animal, "them" is plural pronoun, 30 couple hounds ran after them—hart and [brachet].[2] Human is possible, but I don't get that sense.

Another subject said:

> No strong hypothesis. Maybe something like a buckle (from the sound? brachet sounds like latchet?) Something on a harness worn by the hart?[3]

So far, so little. If you never see the word again, it probably doesn't matter whether you can figure out its meaning, or whether the meaning you figure out is correct.

But suppose that you continue reading, and come across this passage:

(2) … as he [the hart] went by the sideboard, the white **brachet** bit him …. (p. 66.)

Can you now say a bit more about what 'brachet' means? Subject 1 now said:

> Biting in the buttock and pulling out piece implies a sharp toothed animal, not human.

Subject 2 said:

> Maybe … [an] animal. It bites the hart, but that might be figurative—a pin or a buckle might "bite".

Continuing on,[4] our subjects hypothesized as follows:

(3) … the knight arose, took up the **brachet**, … and rode away with the **brachet**. (p. 66.)

(4) … a lady came in … and cried aloud to King Arthur, 'Sire, … the **brachet** is mine …. (p. 66.)

(10) … there was the white **brachet** which bayed at him fast. (p. 72.)

[2] The nonsense word 'arigosin' replaced the word 'brachet' for this subject, since he already knew the term 'brachet'.
[3] The full protocols for 'brachet' and other words are in Ehrlich 1995.
[4] Here, we omit passages that do not contain the word 'brachet'.

At this point, if you already believed that an animal that bayed was probably a dog, you could be pretty sure that a brachet was a dog. Subject 1, in fact, said:

> Only dogs and wolves bay, and a lady would not keep a wolf. This is definitely a dog.

Subject 2, a bit more cautious perhaps, said:

> Confirmation that a brachet is an animal: it "bayed at him fast". Maybe a kind of dog? (Dogs bay). Discard buckle hypothesis.

If you next read this:

(18) … the hart lay dead …; a **brachet** was biting on his throat, and other hounds came behind. (p. 86.)

you can be sure that a brachet is a dog—indeed, a hound (or hunting dog)—because of the phrase 'other hounds'. You have now got a pretty good hypothesis about what 'brachet' means. Subject 1's final definition was that a brachet is "a breed of hunting dog"; subject 2's final definition was that a brachet is "a hound, a hunting dog".

Let's see how our system, named 'Cassie', fares. Passage (1), above, is too hard for her to understand, in part because the archaic English is beyond her grammatical capacities (and, in any case, we are not interested in writing a grammar for this dialect of English), and in part because, in fact, Cassie uses as input a formal-language version of a simplified English version of this passage. (The formal-language version is the intended output (i.e., the semantic interpretation, in SNePS) of a natural-language-understanding algorithm whose input will be the simplified English version.) The simplified English version of passage (1) is:

(1S) ^{1}A hart runs into King Arthur's hall. ^{2}A white brachet is next to the hart.

The formal-language versions of (1S^1) and (1S^2) consist of the following 7 propositions (the expressions of the form `B`*n* are node labels in the SNePS semantic-network representations of these sentences, which can be thought of as Skolem constants; for more details, see, e.g., Rapaport, Shapiro, & Wiebe 1997, §3.1):

(1S^{1}a) In the story, `B17` is a hart.

(1S^{1}b) In the story, `B17` runs into `B18`.

(1S^{1}c) In the story, `B18` is a hall.

(1S^{1}d) In the story, the hall (`B18`) is a hall of its possessor, King Arthur (`B3`).

(1S^{2}a) In the story, `B19` is a brachet.

(1S^{2}b) In the story, the brachet has the property "white".

(1S^{2}c) Brachets are a subclass of physical object.

Proposition (1S^{2}c) is not part of the input, but is *deduced* by Cassie, since Cassie believes that only physical objects have color. (The hedge, "in the story", will be explained later—it is the "knowledge category" *story* (§5); see Rapaport 1991 and Rapaport & Shapiro 1995 for some relevant discussion.)

At this point, we ask Cassie to define 'brachet':

```
--> (defn_noun 'brachet)

((CLASS INCLUSION    = (phys obj))
 STRUCTURE           = NIL
 FUNCTION            = NIL
 ACTIONS             = (NIL)
 OWNERSHIP           = NIL
 POSSIBLE PROPERTIES = ((white))
 SYNONYMS            = NIL)
```

This definition can be paraphrased as: "A brachet is a physical object that may be white".

Omitting the formal-language versions, our demo of Cassie trying to figure out what 'brachet' means continues as follows:

… The **brachet** bites the hart's buttock.

```
--> (defn_noun 'brachet)

((CLASS INCLUSION    = (animal))
 STRUCTURE           = NIL
 FUNCTION            = NIL
 ACTIONS =
   ((POSSIBLE ACTIONS = (bite)))
 OWNERSHIP           = NIL
 POSSIBLE PROPERTIES = ((small white))
 SYNONYMS            = NIL)
```

This can be paraphrased as "A brachet is an animal that may bite and can be small and white".

… The lady says that she wants the **brachet**.

```
--> (defn_noun 'brachet)

((CLASS INCLUSION    = (animal))
 STRUCTURE           = NIL
 FUNCTION            = NIL
 ACTIONS =
   ((POSSIBLE ACTIONS = (bite)))
 OWNERSHIP           = NIL
 POSSIBLE PROPERTIES = ((small valuable white))
 SYNONYMS            = NIL)
```

This can be paraphrased as "A brachet is an animal that may bite and can be small, valuable, and white".

… The **brachet** bays in the direction of Sir Tor.

```
--> (defn_noun 'brachet)

(A BRACHET IS A KIND OF (dog)
 ACTIONS =
   ((POSSIBLE ACTIONS = (bay bite)))
 FUNCTION            = ((hunt))
 STRUCTURE           = NIL
 OWNERSHIP           = NIL
 SYNONYMS            = NIL)
```

This final definition can be paraphrased as "A brachet is a hunting dog that can bay and bite".

There are two important things to note about this demo. First, Cassie's behavior is very similar to those of the two human subjects whose protocols we described above. Second, all three natural-language-understanding systems (human and artificial) converged on roughly the same definition, which is not all that far from that of the *Oxford English Dictionary*, according to which a brachet (or 'brach') is "a kind of hound which hunts by scent" (*Compact Edition of the OED*, Vol. 1, p. 261 (= Vol. 1, p. 1043 of the original edition)). Granted, neither Cassie nor the humans included "scent" in their definitions; however, scent was not mentioned in the Malory text, so there was no reason they should have included it, nor is it necessary for understanding the text.

3 Fundamental Theses

We hold that linguistic contexts (and their mental counterparts) can provide meanings for expressions (Rapaport 1981). This includes non-referential expressions: If 'unicorn' is an unknown term, and we read that a unicorn looks like a horse with one horn, then we can hypothesize a meaning for 'unicorn'—e.g., a one-horned, horselike animal—even though the term 'unicorn' does not refer (cf. Meinong 1910: 25f, Rapaport 1981). A reader understands a narrative (or other) text by interpreting the sentences of the text. The reader's interpretation is a mapping from the sentences (considered as a syntactic domain) to the reader's mental concepts (considered as the semantic domain). (In this way, semantics arises out of syntax, and Searle's Chinese-Room Argument (Searle 1980) can be refuted; see Rapaport 1988, 1995.) We can then take the meaning of a word (as understood by the reader—i.e., a cognitive agent) to be its position in a network of words, propositions, and other concepts (Quillian 1968, 1969). This gives rise to a conceptual-role semantics (see, e.g., Sellars 1955 [1963]; and Harman 1974, 1982 inter alia; see Fodor & Lepore 1992 for objections).

Most importantly, we claim that a meaning for a word *can* be determined from any context, can be *revised* and refined upon further encounters with it, and "*converges*" to a dictionary-like definition given enough context and exposures to it. The context can be minimal, or even empty. Consider the sentence "Tommy broke a **weti**", with unknown word 'weti'. With some reasonable, but minimal, background knowledge, we might theorize that a weti is a breakable physical object. But even with no background knowledge or other contextual information (other than grammatical structure), we could theorize that a weti is something that Tommy broke, by "solving" the sentence for its unknown term, as in algebra (cf. Higginbotham 1989).

Each encounter with the unknown word yields a definition—a hypothesis about its meaning. Of great importance, subsequent encounters provide opportunities for *unsupervised* revision of this hypothesis, with no (human) "trainers" or "error-correction" techniques. The hypothesized definitions are *not* guaranteed to converge to a "correct" meaning (if such exists) but to one stable with respect to further encounters. Finally, no *domain-specific* background information is required for developing the definition: The system need not be an expert in what it's reading.

Clearly, however, the more background knowledge it has (and the larger the textual context), the better the hypothesized definition will be. This suggests that there are two kinds of meaning that we need to consider. In an idiolectic sense, the meaning of a word for a cognitive agent is determined by idiosyncratic experience with it. The contextual meaning described above includes a word's relation to every concept in the agent's mind. Thus, the extreme interpretation of "meaning as context" defines every word in terms of every other word an agent knows. This holistic kind of meaning is circular and too unwieldy for use. In another sense, the meaning of a word is its dictionary definition, usually containing less information. Thus, we limit the connections used for the definition by selecting particular *kinds* of information. Not all concepts within a given subnetwork are equally salient to a dictionary-style definition of a word. People abstract certain conventional information about words to use as a definition.

Two features of our system mesh nicely with these desiderata, summarized as the advantages of learning over being told: (1) Being told requires human intervention. *Our system operates independently of a human teacher or trainer* (with one eliminable exception). (2) Learning is necessary, since one can't predict all information needed to understand unconstrained, domain-independent text. *Our system does not constrain the subject matter ("domain") of the text.* Although we are primarily concerned with narrative text, our techniques are general. Given an appropriate grammar, our algorithms produce domain-independent definitions, albeit ones dependent on the system's background knowledge: The more background knowledge it has, the better its definitions will be, and the more quickly they will "converge". The system does *not* develop "correct" definitions, but *dictionary-like* definitions enabling it to continue understanding the text.

4 Psychological Evidence

Our theory is informed and supported by psychological research on how humans store and access word meanings and expand their vocabularies once the basics of language have been acquired.

4.1 Johnson-Laird

Philip N. Johnson-Laird's (1987) theory about the mental representation of words asserts that understanding a word (and a sentence containing a word) does not imply that one has a readily accessible definition of that word stored in one's mind. Various aspects of a word's meaning may be called to mind by sentences for which those aspects are relevant, without calling the entire definition to mind. He describes experiments showing that responses to questions about specific aspects of a word's meaning come faster following a priming sentence that calls that aspect to mind than to questions where the preceding sentence uses the word, but does not call that aspect to mind. The same speed-up occurs when the priming is a result of factual inference about a word's referent as when it results from selectional restriction on the word's sense.

Linguistic context has at least three effects on the interpretation of words. It can enable selection of an appropriate sense of a truly ambiguous word. It can lead to inference of a more specific referent than is strictly warranted by the meaning of an unambiguous word. It can call to mind particular aspects of a word's meaning at the expense of other aspects. In each case, the mental representation of real or imaginary referents is important to the understanding of the word as used.

Some aspects of a word's meaning are more salient to understanding its use than others. Since the evidence indicates that we do not retrieve definitions in their entirety when understanding sentences, we may not notice gaps in our lexical knowledge, so long as we can retrieve the aspects of meaning necessary to understanding the sentence. Such gaps point to the importance of the acquisition process. One can learn a word's meaning by being told that meaning, or one can infer its meaning from encountering it in use.

Even fairly early in childhood, learning word meanings may be as much a matter of making inferences from linguistic context as of simple association. Johnson-Laird (1987) reports on experiments in which 3- and 4-year-old children listened to stories involving a novel verb. From hearing the verb used in sentences that contained other words they already understood, the children were able to perform selectional restriction on the arguments to the new verb. Children have also been shown to be able to learn aspects of the meanings of nonsense nouns from hearing them used in the context of familiar verbs.

Johnson-Laird suggests that lexical learning involves a sort of "bootstrapping" in which, once a fragment of language has been mapped onto a child's internal representation of states of affairs in the world, other words are acquired either indirectly from context or directly from explicit definition. Words may also be of mixed acquisition; different individuals will acquire a given word differently. Some words can be completely lexically specified; others (e.g., natural-kind terms) cannot, but instead rely in part on a schema of default information based on a prototypical exemplar.

Johnson-Laird (1987: 579) summarizes parts of his theory of lexical meanings as follows: (1) Comprehension requires the listener to construct a model of the state of affairs described by the discourse. (2) There is a mental dictionary that contains entries in which the senses of words are represented. (But cf. Johnson-Laird 1987: 563.) (3) A lexical entry may be incomplete as a result of ignorance or because the word is a theoretical term with an intrinsically incomplete sense. (4) The senses of words can be acquired from definitions or from encountering instances of the word in use. (5) Corresponding to the method of acquisition, elements of a lexical representation can consist of (a) relations to other words, which could be represented by a mechanism akin to a semantic network, and (b) ineffable primitives that are used in constructing and manipulating mental models of the world.

Cassie understands narrative input by building mental representations of the information contained in the narrative; forming concepts of individuals, propositions, and events described; and connecting them with her prior knowledge. Her understanding of a concept in narrative is precisely that concept's connections to the rest of the narrative, together with its connections (if any) to previously acquired knowledge.

We adopt the idea that lexical entries have aspects of meaning connected to them in a semantic network, but do not have compiled, dictionary-style definitions permanently attached. Cassie selects salient features from her knowledge of a word when asked to define it, but does not permanently store those features as a definition. Our semantic network, however, includes meaning postulates (represented as rules) that Cassie can use as part (but not all) of her knowledge for producing definitions. Some information (e.g., about natural kinds) is expressed as default information. Thus, our approach is compatible with Johnson-Laird's theory and experiments.

4.2 Elshout-Mohr and van Daalen-Kapteijns

Marianne Elshout-Mohr and Maartje M. van Daalen-Kapteijns (1987) treat verbal comprehension as a mental skill involving procedural and propositional knowledge. They hold that it is useful to have a "meaning unit" that is stable across contexts, so that new contexts may provide new information at the same time that an established understanding of a word allows interpretation in a new context.

In one experiment, they chose students with high and low verbal skills as measured by standard tests. Subjects were asked to think aloud as they tried to determine from context the meanings of invented words. The neologisms filled lexical gaps, so that they would refer to things with which the subjects were familiar, but would not have direct synonyms that could be substituted for them. (For example, 'kolper': a window that transmits little light because of something outside it.) A series of sentences were presented, one per page, in which the new word was used: The first roughly indicated its superordinate:

> When you are used to a broad view it is quite depressing when you come to live in a room with one or two **kolpers** fronting on a courtyard.

The second and third distinguished it from other concepts in the same superordinate class:

> He virtually always studied in the library, as at home he had to work by artificial light all day because of those **kolpers**.

> During a heat wave a lot of people all of a sudden want to have **kolpers**, so the sales of sunblinds then reach a peak.

The fourth and fifth provided counterexamples:

> I was afraid the room might have **kolpers**, but when I went and saw it, it turned out that plenty of sunlight came into it.

> In those houses, you're stuck with **kolpers** all summer, but fortunately once the leaves have fallen off that isn't so any more."

Subjects were aware of the need to construct a definition rather than search for a synonym. They were asked to report on new information gained from each sentence without reviewing previous pages. It was thought that this would tax their working memory, which was considered important, because most words learned from context are not learned in situations where word acquisition is the primary focus of cognition. Rather, one usually reads for the content of a text and acquires new vocabulary incidentally (if at all).

Most subjects tried to compare the unknown word with at least one familiar word. Those with high verbal skills used the model analytically, as a group of separable components that could be individually compared with further information. New information compatible with all facets of the model could be added; conflicting facets of the model could be replaced with new information. Those with lower verbal skills tended to use the model holistically, with new compatible information broadening or restricting the domain (Kiersey 1982) and incompatible information causing the rejection of the model (and perhaps the adoption of a different model).

According to the authors, in the task of word learning, a model (1) provides a plan for knowledge retrieval; all aspects of the semantic unit of the model are accessible; (2) provides a frame-like structure for the meaning to be acquired, with certain slots to fill; (3) allows conventional aspects of definitions to steer abstraction toward similar conventions in the new definition; and (4) provides default information to fill slots in the anticipated structure. The meaning of the new word is presumed to inherit aspects of meaning connected with the model unless otherwise specified.

Cassie uses her prior knowledge and the network that represents her understanding of the story (up to and including the sentence containing that word) to establish a definitional framework for the target word. This framework is not an analogical model in Elshout-Mohr and van Daalen-Kapteijns's sense, but does provide a plan for knowledge retrieval and a structure with certain slots to fill. The framework for nouns includes slots for synonyms and for hypernyms from which the target word can inherit aspects of meaning.

Because we wish our system to model an individual of high verbal ability (for general usefulness and specifically for use as a lexicographer's assistant), it does use its selected framework analytically, although new information may cause the system to select a different framework than that chosen after the first encounter with the target word.

Elshout-Mohr and van Daalen-Kapteijns suggest that children's verbal skills be developed by instruction in distinguishing between idiosyncratic experiences with a word and the more general experiences associated therewith, and in constraining the richness of individual experience by selecting a limited number of aspects of meaning. Especially important is the ability to select structural and functional aspects of a concept that are "in common with" and "in distinction of" other known concepts. Developing theory and technique for such movement from idiosyncratic understanding to conventional dictionary definitions is a central portion of our research.

4.3 Sternberg

Robert Sternberg (1987: 91) gives the following example:

> Although for the others the party was a splendid success, the couple there on the blind date was not enjoying the festivities in the least. An **acapnotic**, he disliked her smoking; and when he removed his hat, she, who preferred "ageless" men, eyed his increasing **phalacrosis** and grimaced.

It is easy to guess what 'acapnotic' and 'phalacrosis' might mean (although the latter is ambiguous; does it mean "baldness" or "grey hair"?). Sternberg holds that three processes are applied in acquiring words from context: (a) distinguishing irrelevant from relevant information, (b) selectively combining relevant clues, and (c) comparing what has been selected and combined with previous knowledge. These processes operate on a basis of several types of cues: (1) temporal cues regarding the duration or frequency of X, or when X can occur; (2) spatial cues regarding the location of X or possible locations where X can sometimes be found; (3) value cues regarding the worth or desirability of X, or the kinds of affects X arouses; (4) stative descriptive cues regarding the properties of X (size, color, etc.); (5) functional descriptive cues regarding possible purposes of X, actions X can perform, or potential uses of X; (6) cues regarding the possible causes of, or enabling conditions for, X; (7) cues regarding one or more classes to which X belongs, or other members of one or more classes of which X is a member; (8) equivalence cues regarding the meaning of X, or contrast (e.g., antonymy) to the meaning of X.

Sternberg's experiments show that readers who are trained in his three processes of recognizing relevant cues, combining cues, and comparing with known terms do better at defining new words than those who receive training only in the eight types of cues available, although any training in cue types produces better results than word memorization or no training.

Our system is designed to select certain information as most relevant, if it is present. For example, we follow Johnson-Laird as well as Elshout-Mohr and van Daalen-Kapteijns in emphasizing the structural and functional information about physical objects. If, however, such information is lacking, our system uses such other cue types as may be available. We also combine relevant information, and compare new words with known words in certain specific ways, such as possible synonymy or hyponymy.

5 Implementation

Our system, "Cassie", consists of the SNePS-2.1 knowledge-representation and reasoning system (Shapiro 1979; Shapiro & Rapaport 1987, 1992, 1995), and a knowledge base representing Cassie's background knowledge. Currently, the knowledge base is hand-coded, since *how* she acquired this knowledge is irrelevant. Although we begin with a "toy" knowledge base, each of our tests includes all previous information, so the knowledge base grows as we test more words. Cassie's input consists of (1) information from the text being read and (2) questions that trigger a deductive search of the knowledge base (e.g., "What does ⟨word⟩ mean?"). Output consists of a report of Cassie's current definition of the word, or answers to other queries.

SNePS has an English lexicon, morphological analyzer/synthesizer, and a generalized augmented-transition-network parser-generator that translates English input directly into a propositional semantic network without building an intermediate parse tree (Shapiro 1982, 1989; Rapaport 1988, 1991; Shapiro & Rapaport 1995). All information, including propositions, is represented by SNePS nodes; propositions about propositions can also be represented. Labeled arcs form the underlying syntactic structure of SNePS, embodied in the restriction that one cannot add an arc between two existing nodes, which would be tantamount to a proposition not represented by a node. Arc-paths can be defined for path-based inference, including property inheritance. Nodes and represented concepts are in 1–1 correspondence; this uniqueness principle guarantees that nodes are shared whenever possible and that nodes represent intensional objects (e.g., concepts, propositions, properties, and objects of thought including fictional entities, non-existents, and impossible objects; Shapiro & Rapaport 1987, 1991).

SNePS's inference package accepts rules for deductive and default reasoning, allowing Cassie to infer "probable" conclusions in the absence of contrary information. When combinations of asserted propositions lead to a contradiction, SNeBR, the SNePS belief-revision package, allows Cassie to remove from the inconsistent context one or more of those propositions (Martins & Shapiro 1988). Once a premise is no longer asserted, the conclusions that depended on it are no longer asserted in that context. Cassie uses SNeBR and SNePSwD, a default belief-revision system that enables automatic revision (Cravo & Martins 1993; Martins & Cravo 1991), to revise beliefs about the meanings of words.

Because we wish our system to be able to revise its beliefs with a minimum of human interference, we have partially automated the selection and revision of the culprit node (i.e., the erroneous assertion that led to the contradiction); we are exploring techniques for full automation. To facilitate this process, we tag each of the asserted propositions in the knowledge base with a knowledge category. Maria Cravo and João P. Martins (1993) have developed a utility for creating orderings of propositions within the network. We use this ordering utility to build a hierarchy of certainty in which all propositions with a particular knowledge category are given higher priority than those whose knowledge category indicates less certainty of belief. Then, when SNeBR is invoked by a derived contradiction, the belief(s) with least priority (i.e., held with least certainty) in the conflict set will be selected for revision. As originally designed, once the culprit belief had been deleted, SNeBR would ask the user if she wished to add any new propositions to the network. We have modified SNeBR so that it now automatically creates a revision of the culprit belief (see Ehrlich 1995 for the algorithms).

The current version of SNePS allows ordinary deductive reasoning. SNePSwD also allows "default reasoning" (Cravo & Martins 1993). Default rules allow the system to infer "probable" conclusions in the absence of specific information to the contrary. For example, if the system knows that, by default, birds fly, but also knows that penguins do not fly and that Opus is a penguin, then, despite inferring that Opus is a bird, the system will not conclude that Opus flies, because the default rule is not applicable to Opus. This works well when all relevant information is known before the attempt to deduce whether Opus flies. However, if the system must make a decision before learning that penguins are flightless, it will need to revise that decision once more information is gained.

Using SNeBR for revising erroneous beliefs is not appropriate in all situations. Rather than use default rules that rely on being told in advance whether they are applicable in certain cases, we employ a method, based on theoretical work by J. Terry Nutter (1983), in which some rules have consequents marked as presumably true. Frequently, this avoids the need for non-monotonicity. In our example above, we would have the rule that, if something is a bird, then presumably it flies. Opus the penguin, being a kind of bird, would fit the antecedent, and we would conclude that presumably Opus flies. Learning that penguins do not fly does not, then, produce a contradiction, but rather the concatenation that Opus could be presumed to fly though in fact he does not.

When a human reader encounters a discrepancy between the way a word is used in text and his previous understanding of that word, he must either assume that the word is used incorrectly or decide that his previous understanding requires revision. When our system encounters a contradiction derived from combining story information with background knowledge, it must decide which of the premises leading to the contradiction should be revised (cf. Rapaport 1991, Rapaport & Shapiro 1995). To facilitate such decisions, each of the assertions that we build into the system's knowledge base and each of the assertions in the story is tagged with a knowledge category (`kn_cat`). Assertions having no `kn_cat` attached are beliefs that the system has derived. (SNeBR assigns each proposition an "origin tag" as either a hypothesis (a proposition received as input) or a derivation. If the proposition is derived, information about the hypotheses used in the derivation is also stored (Martins & Shapiro 1988). Our `kn_cat` tags may be considered complementary to SNeBR's origin tags, though implemented differently.) These categories are ordered in a hierarchy of certainty of belief, so that the system can restrict the field from which it chooses a belief for revision to those premises believed with the least certainty. Ideally, there would be only one such premise, but if there are more, then other means must be used to select among them (see below). Following is a description of the hierarchy of `kn_cats` ranged from greatest certainty of belief to least:

1. `kn_cat` *intrinsic*: Essentially, facts about language, including simple assertions, as well as some rules; background, or "world", knowledge of a very basic or fundamental sort. Intrinsic facts are found in the knowledge base, not in stories (at least, usually not). For example: The temporal relation "before" is transitive; containment of an item in a class implies containment of that item in superclasses of that class; encountering the usage ⟨verb⟩(agent, object, indobj) implies that ⟨verb⟩ can be bitransitive.

2. `kn_cat` *story*: Information present in the story being read, including stated propositions and propositions implicit in the sentence (necessary for parsing the sentence); the SNePS representation that would be built on parsing a sentence in the story. For example, in the sentence "Sir Gryflette left his house and rode to town", we have the following story facts: Someone is named Sir Gryflette. That someone left his house. That someone rode to town. (Other examples are ($1S^1$a)–($1S^2$b), above.)

3. `kn_cat` *life*: Background knowledge expressed as simple assertions without variables or inference. Examples of

life facts include taxonomies (e.g., dogs are a subclass of animals) and assertions about individuals (e.g., Merlin is a wizard. Bill Clinton is a Democrat.).

4. `kn_cat` *story-comp*: Information not directly present in the story, but inferred by the reader to make sense of the story. This is based on Erwin M. Segal's concept of "story completion" (Rapaport et al. 1989a,b; Segal 1995). Story completion uses background knowledge, but isn't the background knowledge itself. Few (if any) assertions should be tagged with this `kn_cat`, since any necessary story completion should (ideally) be derived by the system. We include it here to cover cases where a gap in the system's knowledge base might leave it unable to infer some fact necessary to understanding the story. Using the example from the category of story facts, story completion facts might include: Sir Gryflette is a knight; Sir Gryflette mounted his horse between leaving his house and riding to town.

5. `kn_cat` *life-rule.1*: Background knowledge represented as rules for inference, using variables; rules reflecting common, everyday knowledge. For example: If x bears young, then x is a mammal; if x is a weapon, then the function of x is to do damage; if x dresses y, then y wears clothing.

6. `kn_cat` *life-rule.2*: Background knowledge represented as rules for inference, using variables, that rely on specialized, non-everyday information. For example: if x smites y, then x kills y by hitting y.

7. `kn_cat` *questionable*: A rule that has already been subjected to revision because its original form led to a contradiction. For example: If x smites y, then x hits y and possibly kills y. This is the only kn_cat that is never a part of input. The system attaches this tag when it revises a rule that was tagged as a *life-rule.2*. It is intended as a temporary classification for use while the system looks for confirmation of its revision. Once the system settles on a particular revision, the revised rule is tagged as a *life-rule.2*.

In case of contradiction, our system selects, from among the conflicting propositions, a proposition of greatest uncertainty as a candidate for revision. If the highest level of uncertainty present in the conflict set occurs in only one belief, that belief will be revised. If several alternatives exist with the same (highest present) `kn_cat`, then the system selects for the presence of a verb in the antecedent, since it seems that humans more readily revise beliefs about verbs than about nouns (Gentner 1981). If this is still insufficient to yield a single culprit, then, in the current implementation, a human user must decide among any remaining alternatives. Ideally, the system would use discourse information to make the decision between possible culprits at the same level of certainty. For example, in the case of combatants dressing shields and spears before fighting (see §7.2, below), the rule about what it means to dress something might be selected for revision because it is unrelated to the topic of fighting, whereas shields and spears are closely associated with the topic.

We have described a belief hierarchy of seven levels. In practice, it seems that four would probably suffice. We could have handled all our example revisions without separating the non-rule information into *intrinsic*, *story*, *life*, and *story-comp*letion information (though some intrinsic facts expressed as rules would have had to be expressed differently). A hierarchy consisting of non-rules, entrenched rules (*life-rule.1*), non-entrenched rules (*life-rule.2*), and rules under revision (*questionable*) would be enough to cover the cases with which we have dealt, since we have never had to revise a non-rule belief. However, it is at least possible that a contradiction might arise between two or more beliefs, none of which are rules.

As mentioned above, story-completion information is something of a special case, and will usually be derived; that is, it will be a conclusion based on background information and story information. When SNeBR detects a contradiction, it assembles a conflict set consisting of those *premises* that led to the contradiction. Derived conclusions are not part of the conflict set. Therefore, except where story-completion information is directly input, we need not be concerned about selecting it as the culprit whose error led to a contradiction.

So, if we were to encounter a contradiction whose conflict set contained no rules, we would need some method for selecting a culprit from among various non-rule facts. Our approach assumes that words in the story are used correctly and that information presented in the story is true, at least within the context of the story. Therefore, it makes sense to distinguish between story facts and background facts. But not all background facts are equally entrenched. Some, such as the examples of `kn_cat` *intrinsic* given above, seem so basic to our understanding that they should be immune from revision (even if they happen to be expressed as rules). Others seem less basic, and may be revised if they conflict with story facts. The distinction between assertions tagged *intrinsic* and those tagged *life* is not exactly analogous to distinctions between necessary and contingent properties, or between analytic and synthetic definitions, but a loose analogy to such distinctions may capture some sense of the distinction being drawn here (cf. Rapaport 1991, Rapaport & Shapiro 1995).

At present, all the `kn_cats` (except for *questionable*) are assigned by a human at the time the proposition is input. The assignment of the `kn_cat` *story* could be handled automatically: The system would simply include it as a part of each proposition built from the parse of a sentence in a story (as in Rapaport 1991, Rapaport & Shapiro 1995). Since *story-comp* is only a stop-gap measure, we need not worry about how the system might assign it: Either it wouldn't make any such assignment, or it would tag all derived propositions as being derived. (The latter is already done by SNeBR, but invisibly; cf. Martins & Shapiro 1988.) This leaves us with the question of how the system might categorize the non-derived assertions in its knowledge base. Rules can be readily distinguished from non-rules, so the question breaks down into two parts: How do we tell an entrenched rule (*life-rule.1*) from a less-entrenched rule (*life-rule.2*), and how do we tell an entrenched fact (*intrinsic*) from a less-entrenched fact (*life*)? We make the distinction based on an intuitive feeling for how basic a concept is, or how familiar we are with a concept. How such intuitions can be formalized and automated is an open question. We may have to continue for a while to tell Cassie how strongly she holds her beliefs.

Once a contradiction is detected and a culprit proposition selected, if it is a rule (as it usually is), then it may be necessary to select which of several consequents actually produced the contradiction. The system checks each consequent against the justification support (the premises from which a proposition is derived) of the deduced propositions that are in direct conflict with each other to find the culprit consequent within the culprit rule (i.e., the rule that has been determined to be in error).

6 Algorithms

Our algorithms hypothesize and revise meanings for nouns and verbs that are unknown, misunderstood, or being used in a new way. Applying the principle that the meaning of a term is its location in the network of background and story information, our algorithms deductively search the network for information appropriate to a dictionary-like definition, assuming our grammar has identified the unknown word as a noun or a verb. The algorithms (Ehrlich 1995) are shown in Appendices 1 and 2, and are illustrated by example here.

Cassie was provided with background information for understanding King Arthur stories (Malory 1470). As we saw above, when presented with a sequence of passages containing the unknown noun 'brachet', Cassie developed a theory that a brachet was a dog whose function is to hunt and that can bay and bite. However, based on the first context in which the term appeared ("… there came a white hart running into the hall with a white brachet next to him, …"), her initial hypothesis was that a brachet was a physical object that may be white. Each time 'brachet' appeared, Cassie was asked to define it. To do so, she deductively searched her background knowledge base, together with the information she had read in the narrative to that point, for information concerning (1) direct class inclusions (especially in a basic-level category), (2) general functions of brachets (in preference to those of individuals), (3) the general structure of brachets (if appropriate, and in preference to those of individuals), (4) acts that brachets perform (partially ordered in terms of universality: probable actions in preference to possible actions, actions attributed to brachets in general in preference to actions of individuals, etc.), (5) possible ownership of brachets, (6) part-whole relationships to other objects, (7) other properties of brachets (when structural and functional description is possible, less salient "other properties" of particular brachets are not reported, although we do report properties that apply to brachets in general), and (8) possible synonyms for 'brachet' (based on similarity of the above attributes). Some of these are based on the psycholinguistic studies of the sort of vocabulary expansion we are modeling (discussed above). In the absence of any of this information, or in the presence of potentially inconsistent information (e.g., if the text says that one brachet hunts and another doesn't), Cassie either leaves certain "slots" in her definitional framework empty, or includes information about particular brachets. Such information is filled in or replaced upon further encounters with the term.

Each query for a definition begins the search from scratch; different information is reported depending on the kind of background information available at the time the query is made. Thus, by querying Cassie after each occurrence of 'brachet', we can see the definition frame "develop" dynamically. However, if we only queried Cassie at the last occurrence of 'brachet', then we would only see the "final" definition frame.

Although the current implementation outputs different definition frames depending on which branch of the algorithm has been taken (see Appendices 1 and 2), a "unified" definition frame could be used, as shown in Table 1. Each row is a slot. Each column represents a kind of term; e.g., (1) a term that is a basic-level category, such as 'dog', (2) a term that is a subclass of a basic-level category, such as 'brachet', etc. (See Rosch 1978 for the notion of "basic-

Slot	Basic-Level Category	Subclass of Basic-Level Category	Subclass of Animal	Subclass of Physical Object	Subclass of Abstract Object	No Class Inclusions
actions	√	√	√	√	√	√
ownership	√	√	√	√	√	√
function	√	√	√	√	√	√
structure	√	√	√	√	⊗	√
immediate superclass	√	√	√	√	√	⊗
general stative properties	√	√	√	√	√	⊗
synonyms	√	√	√	√	√	⊗
possible properties	◯	◯	◯	√	√	√
object of act	◯	◯	◯	√	√	√
named individuals	◯	◯	◯	◯	◯	√

Table 1: Unified Definition Frame: √ = reported if known
◯ = not reported (even if known)
⊗ = can't be reported

level categories".) A check-mark (√) means that a slot-filler is reported if it is known. A circle (◯) means that it is not reported, even if it is known, because it is superseded by other slot-fillers. A circled X (⊗) means that that slot cannot be filled (hence, will not be reported) for that kind of term.

7 Other Examples of Vocabulary Acquisition

We have been investigating three types of vocabulary acquisition:

1. *Constructing* a new definition of an *unknown* word (e.g., 'brachet')
2. *Correcting* a definition of a *misunderstood* word (e.g., see the discussion of 'smite', §7.1)
3. *Expanding* the definition of a word being *used in a new sense* (e.g., see the discussion of 'dress', §7.2).

All three can be thought of as revision: (1) "revision" from an empty definition, (2) revision of an incorrect definition, and (3) revision of an incomplete definition. Alternatively, (3) can be thought of as adding (disjoining) a new definition to an already-established one (see §7.2, below).

7.1 'Smite'

Cassie was told that 'to smite' meant "to kill by hitting hard" (a mistaken belief actually held by one of us (Rapaport) before reading Malory 1470). Passages in which various characters were smitten but then continued to act triggered SNeBR, which asks Cassie which of several possible "culprit" propositions in the knowledge base to remove in order to block inconsistencies. Ideally, Cassie then decides which belief to revise. A set of rules for replacing discarded definitions with revised definitions is being developed. E.g., suppose the culprit were:

If x smites y, then: x hits y & y is dead & x hitting y causes y to be dead.

On first encountering a smitee who survives, substitute these rules:

1. If x smites y, then: x hits y & possibly y is dead;
2. If x smites y & y is dead, then x hitting y causes y to be dead.

If asked for a definition of 'smite' now, Cassie will report that the result of smiting is that x hits y and possibly y is dead. The only human intervention is to tell Cassie to order her beliefs (she does the ordering, based on the knowledge categories, but has to be nudged to "do it now") and to tell Cassie to *assert* the revised belief she has *already automatically* generated from the lowest-ranked belief in the conflict set.

Here, we sketch Cassie's handling of 'smite', with this background information in the knowledge base:

There is a king named King Arthur.
There is a king named King Lot.
There is a sword named Excalibur.
Excalibur is King Arthur's sword.
Horses are animals.
Kings are persons.
Knights are persons.
Dukes are persons.
"Person" is a basic-level category.
"Horse" is a basic-level category.
"Before" and "after" are transitive relations.
If x is dead at time t, x can perform no actions at t or at any subsequent time.
If x belongs to a subclass of person, x is a person.
If a person acts, the act performed is an action.
If an agent acts on an object, and there is an indirect object of the action, then the action is bitransitive.
If an agent acts on an object, then the action is transitive.
If an agent acts on itself, then the action is reflexive.
If x is hurt at time t, then x is not dead at t.
If x is not dead at time t, then x was not dead at any prior time.
If x smites y at time t, then x hits y at t, and y is dead at t, and the hitting caused the death.

Note that the last is the only information about 'smite' in the knowledge base.

Cassie is then given a sequence of passages containing 'smite' (adapted from Malory 1470: 13ff), interspersed with questions and requests for definitions:

Passage S1: King Arthur turned himself and his horse. He smote before and behind. His horse was slain. King Lot smote down King Arthur.

Definition S1: A person can smite a person. If x smites y at time t, then x hits y at t, and y is dead at t.

Question S1: What properties does King Arthur have?

Reply S1: King Arthur is dead.

Passage S2: King Arthur's knights rescued him. They sat him on a horse. He drew Excalibur.

Question S2: When did King Arthur draw [Excalibur]?

The inference required to reply to **Question S2** triggers SNeBR, which reports that King Arthur's drawing (i.e., acting) is inconsistent with his being dead. Cassie automatically removes the proposition reporting her belief that smiting entails killing, which is replaced with two beliefs: that although smiting entails hitting, it only possibly entails killing, and that if smiting results in a death, then the hitting is the cause of death. These rules are *not* built in; they are *inferred* by revision rules. This results in:

Definition S2: A person can smite a person. If x smites y at time t, then x hits y at t and possibly y is dead at t.

Passage S3: Two of King Claudas's knights rode toward a passage. Sir Ulfyas and Sir Brastias rode ahead. Sir Ulfyas smote down one of King Claudas's two knights. Sir Brastias smote down the other knight. Sir Ulfyas and Sir Brastias rode ahead. Sir Ulfyas fought and unhorsed another of Claudas's knights. Sir Brastias fought and unhorsed the last of Claudas's knights. Sir Ulfyas and Sir Brastias laid King Claudas's last two knights on the ground. All of King Claudas's knights were hurt and bruised.

The information that the knights were hurt was added in forward-chaining mode to allow Cassie to notice that that they were still alive at the time they were hurt and therefore could not have died earlier at the time they were smitten. Cassie has now heard of two cases in a row (King Arthur, and the two knights) where a smitee has survived being smitten, with no intervening cases of death by smiting, yielding:

Definition S3: A person can smite a person. If x smites y at time t, then x hits y at t.

Further encounters with 'smite' cause no further revisions. The definition has stabilized ("converged") in a manner similar to the human protocols on the same passages.

7.2 'Dress'

A third case is exemplified by 'dress', which Cassie antecedently understood to mean "put clothes on (something)", a well-entrenched meaning that should *not* be rejected. Thus, her background knowledge for this example included the following two beliefs whose `kn_cat` = *life-rule.1*.

1. dresses$(x, y) \Rightarrow \exists z$[clothing$(z)$ & wears(y, z)].
2. Spears don't wear clothing.

Again, Cassie was given a sequence of passages containing 'dress' (adapted from Malory 1470) interspersed with questions and requests for definitions:

Passage D1 King Arthur dressed himself.

Definition D1 A person can dress itself; result: it wears clothing.

Passage D2 King Claudas dressed his spear ....

At this point, Cassie infers that King Claudas's spear wears clothing.

Question D2 What wears clothing?

This question invokes SNeBR, which detects a contradiction and automatically replaces background belief (1), *not background belief (2)*, because of the occurrence of a verb in the antecedent, following Gentner 1981 (see §5, above). There follow several passages in the text in which dressing spears precedes fighting. Rather than *rejecting* the prior definition, we *add* to it. Cassie decides that to dress is *either* to put clothes on *or* to prepare for battle:

Definition D3 A person can dress a spear or a person; result: the person wears clothing or the person is enabled to fight.

Admittedly, this is not perfectly satisfactory. Among the improvements we are planning are a better way of expressing such disjunctive definitions and a method to induce a more general definition. In the present case, further experience with such phrases as 'salad dressing', 'turkey dressing', and so on, might lead one to decide that 'dress' more generally means something like "prepare" (for the day, by putting on clothes; for battle, by preparing one's spear; for eating, by preparing one's salad; for cooking, by preparing one's turkey; and so on).

8 Related Work

We begin with a review of some classic work along these lines, and then turn to more recent work.

8.1 Some classic work

8.1.1 Granger

Richard H. Granger's Foul-Up (1977) was designed specifically to work in conjunction with a Script Applier Mechanism (Schank & Abelson 1977) to allow stories of events to be understood in terms of stereotypical event sequences despite the presence of unknown words. Granger used the syntactic expectations generated by a parser and script-based expectations generated by the currently active script to create a partial definition of an unknown word. For example, if the currently active script is *vehicular accident*, if there is input to the effect that the car struck an elm, and if the word 'elm' is unknown, Foul-Up would deduce that 'elm' is a noun that represents something that can fill the "obstruction" slot in a script for a single-car accident (and therefore must be a physical object).

Once our grammar is in place, we also will use syntactic expectations and morphology to allow the determination of an unknown word's role in a given sentence (e.g., subject, direct object, verb, etc.) as well as its part of speech. We can make deductions similar to Granger's, but in terms of sentences instead of scripts. We also use such background knowledge as is available, even though that knowledge isn't organized into scripts. We would deduce that 'elm' is a noun and that an elm is a thing that can be struck by a car. For us, as for Granger, the single occurrence of the word is not enough to allow any useful inference beyond what is immediately present in the text (although background knowledge might be available about the probable size, location, etc., of objects that a car might be said to strike).

We adopt Granger's use of parsing expectations to begin hypothesizing word definitions. However, the rest of his approach, which involved the use of standard scripts, is less general than our approach. Unknown words do not always occur in the context of a story that fits a standardized script. Furthermore, even if a standardized script may apply to the text overall, the unknown term may not occur as the filler of a standard slot. This does not mean, however, that the context in which the word occurs may not yield useful information in attempting to hypothesize a definition.

Our research also differs from Granger's in that we store the knowledge acquired from reading various texts in a single, general knowledge base, and use this stored knowledge as source of information for definitions. Thus, a definition may be synthesized from several encounters.

8.1.2 Haas and Hendrix

Norman Haas and Gary Hendrix (1983) developed a system in which a tutor specifically gives the system a new concept, and refines it by adding further facts in response to queries from the system as it seeks to relate the new concept to previously acquired knowledge. Although asking questions is a good way to learn, the major focus of our research is to allow the system to learn from context without specific queries or direct instruction. We explore what the system learns when left to read "on its own", and allow it to include only information it can deduce in its definitions. In future work, as already noted in connection with 'dress', we may explore ways in which the system can generate *inductive* hypotheses, which might be used to generate questions to put to a human user, or might simply be kept as questions in the "mind" of our system until more evidence can be found.

8.1.3 Berwick

Another approach to learning verbs (and perhaps other words) was developed by Robert C. Berwick (1983), who attempted to extend the process of analogical learning (as in Patrick Henry Winston's (1975) arch-learning algorithm) to the domain of natural language. Causal networks are created to represent the events in different stories. It is assumed that a word's meaning is known by its position in such a net. A new word, then, is held to be most similar to those known words which play the most similar roles. For example, if murder and assassination both cause their objects to be dead, but love does not cause such an outcome, then 'murder' and 'assassinate' are closer in their meaning than either is to 'love'. Berwick's work, as implemented, required the input of story outlines in specific form, with all relevant causal information explicitly present in the outlines, rather than straight narrative as it occurs in published stories.

While the notion of synonymy (or near synonymy) and similar causes producing similar effects is a useful one, our desire to test our theories on natural text would require that our system infer causal relationships in order to make the sort of comparisons that Berwick's system made. For example, were the system to read that Macbeth assassinated Duncan and then to read that Duncan was dead, it would have to infer that the one fact was the cause of the other. This done, it might use a system similar to Berwick's to infer that 'assassinate' is close to 'murder' in meaning (assuming, of course that it already knew that "*A* murdering *B*" causes *B* to be dead). Unfortunately, giving our system the ability to make such causal inferences is beyond the scope of this project, although given the presence of certain specific causal information in the knowledge base, comparisons between verbs can be made.

We do make use of the notion that words with similar meanings have similar connections in an agent's network of knowledge. The framework we use for defining nouns includes a slot for potential synonyms. The synonym-finding algorithm compares the other definitional information (e.g., class inclusion, structure, and function) that has been discovered about the new word with the same information about a candidate synonym. If they have more in common than in contrast, the system lists them as possible synonyms.

8.2 Some more recent work

8.2.1 Zernik and Dyer

Uri Zernik and Michael G. Dyer (1987) compile definitions of words and figurative phrases from conversation into a hierarchical lexicon, using a pattern constructor that analyzes parsing failures to modify its patterns and a concept constructor that selects a strategy according to background information about the goals and plans a person is likely to have in various situations. If the first interpretation of a phrase is inconsistent with that information, the system queries the user, suggesting various interpretations more consistent with the goals of persons in the story until the user confirms that a correct interpretation has been reached.

However, since we focus on literary, not conversational, discourse, Cassie is not informed by a human user when she misunderstands. As long as the misunderstanding is compatible with further encounters with the word and with her general knowledge, there is no reason for Cassie to revise her understanding. If further reading leads to the conclusion that a previous definition was wrong, Cassie revises her understanding without explicit instruction.

8.2.2 Hastings

Peter M. Hastings (1994; Hastings & Lytinen 1994ab) presents several versions of a system, Camille, that uses knowledge of a given domain to infer a word's meaning. Hastings's approach is like ours: The goal is to allow the system to read and acquire word meanings without the intervention of a human tutor. His approach differs, however, in the types of information sought as the meaning of a word, and in the nature of the knowledge base. For each domain, the initial knowledge base consisted of a taxonomy of relevant objects and actions. Camille attempts to place the unknown word appropriately in the domain hierarchy. To this basic system, Hastings has added: a mutual exclusivity constraint; a script applier allowing Camille to match the unknown word with a known word that has filled the same slot in a particular script, or, for a verb, with a known word whose arguments match those of the target word; and an ability to recognize the existence of multiple senses for a word. In most instances, the meaning sought appears to be synonymy with a known word, unlike Cassie, which can create new concepts (defined in terms of preexisting ones). In one version, however, Camille is given the capacity to create a new node, and insert it into the domain hierarchy. This, however, is only available for unknown nouns. Verbs can be "defined" only by mapping them to their closest synonyms. Hastings's evaluation of Camille's performance is given in terms of "correctness" of word meaning. His focus is on the comparative precision and accuracy of the various versions of Camille as they attempt to map unknown terms onto known nodes. For us, such a notion of "correctness" does not apply.

8.2.3 Siskind

Jeffrey Mark Siskind's (1996) focus is on first-language acquisition during childhood, whereas ours is on mature cognitive agents who already know a large part of their language and are (merely) expanding their vocabulary. Siskind takes as given (1) an utterance, (2) a simultaneous visual perception, (3) a mental representation of the situation perceived, which is caused by it, and (4) an assumption that the utterance means that mental representation. His algorithms assign parts of the mental-representation meaning to parts (words) of the utterance. Given "multiple situations", these algorithms "converge" to "correct word-to-meaning mappings".

Although we also assume an utterance and a mental representation that the utterance means, Cassie does not use visual perception to produce the mental representation. In most cases of reading, any mental representation (including imagery) would be produced by the text, so visual perception of a real-world situation does not arise, except for illustrated texts. Although Cassie does not use illustrations, she could in principle (Srihari & Rapaport 1989, Srihari 1991). Siskind's system begins with a mapping between a whole meaning and a whole utterance, and infers mappings between their parts. Cassie already has both of those mappings and seeks to infer definition-style relations between the unknown word and the rest of the knowledge base. Moreover, it does not make sense in our theory to speak of "correct word-to-meaning mappings". Finally, Siskind claims that his theory provides evidence for "semantic bootstrapping"—using semantics to aid in learning syntax. In contrast, Cassie uses *syntactic* bootstrapping (using syntax to aid in learning semantics; Gleitman 1990, 1994), which seems more reasonable for our situation.

8.2.4 Campbell and Shapiro

A related project is Ontological Mediation, being conducted by Alistair E. Campbell and Stuart C. Shapiro in the SNePS Research Group at SUNY Buffalo. Suppose a speaker uses a word unknown to the listener. An "ontological mediator" agent determines the word's meaning by querying the speaker and listener concerning ontological relations, such as subclass, parts, ownership, or skills, and then defines the word for the listener in terms of the listener's already-known words. (See Campbell & Shapiro 1995, Campbell 1996, and ⟨http://www.cs.buffalo.edu/∼aec/OM/index.html⟩.)

9 Future Work

Much remains to be done: We are now developing and using a grammar for natural-language input and output (Hunt & Koplas 1997), as well as for potential use in determining meanings via morphological (and perhaps etymological) information. We plan to expand our algorithms for verb acquisition, and to investigate adjective acquisition. We need to decide under what conditions to represent the definition in the knowledge base; despite Johnson-Laird's findings, discussed above (§4.1), it is nevertheless true that eventually we do memorize definitions of words after encountering them often enough. We plan to apply our algorithms to proper names: Can we decide who someone is by using our techniques? Finally, to what extent are we developing a formal model of category-definition by exemplars, where we take as input information about a (single) individual, and output a description of a category that that individual falls under?

Appendix 1: Algorithm for Noun Definition

Input unknown noun 'N'.

```
procedure Make-List1 ::=
   list (1) structure of Ns,
        (2) functions of Ns,
        (3) stative properties of Ns only if there are general rules about them.
end; {procedure Make-List1}

procedure Make-List2 ::=
   list (1) direct class inclusions of N,
        (2) actions of Ns that can't be deduced from class inclusions,
        (3) ownership of Ns,
        (4) synonyms of 'N'.
end; {procedure Make-List2}

procedure Make-List3 ::=
   begin
        Make-List2;
        if there is structural or functional info about Ns, then Make-List1
   end
end; {procedure Make-List3}

begin {defn_noun}
   if N represents a basic-level category, then Make-List3
   elsif N represents a subclass of a basic-level category, then
        begin
            report that N is a variety of the basic-level category that includes it;
            if Ns are animals, then list non-redundant acts that Ns perform;
            list if known:
                functions of Ns,
                structural information about Ns,
                ownership of Ns,
                synonyms of 'N';
            list stative properties only if there are general rules about them
        end
   elsif N represents a subclass of animal, then Make-List3
   elsif N represents a subclass of physical object, then
        begin
            Make-List2;
            if system finds structural or functional information about Ns, then Make-List1
            elsif system finds actions of N or synonyms of 'N', then
                begin
                    list them;
                    list possible properties of Ns
                end
            elsif N is an object of an act performed by an agent, then
                begin
                    report that;
                    list possible properties of Ns
                end
        end
```

```
elsif N represents a subclass of abstract object, then
    begin
        list direct class inclusions of N & ownership of Ns;
        if system finds functional information about Ns,
            then list:
                function,
                actions of Ns that can't be deduced from class inclusions,
                stative properties only if there are general rules,
                synonyms for 'N'
            else begin
                    list possible properties of Ns;
                    if system finds actions of N or synonyms for 'N', then list them
                    elsif N is an object of an act performed by an agent, then report that
                end
    end
else {we lack class inclusions, so:}
    begin
        list:
            named individuals of class N,
            ownership,
            possible properties;
        if system finds information on structure, function, actions, then list it
        elsif N is object of act performed by agent, then report that
    end
end.
```

Appendix 2: Algorithm for Verb Definition

Input unknown verb ‘V’.

begin {`defn_verb`}
  report on cause & effect;
  categorize the subject;
  if V is used with an indirect object,
    then categorize objects & indirect object
  elsif V is used with a direct object distinct from its subject,
    then categorize the object
  elsif V is used with its subject as direct object,
    then list the object as “itself”
end.

References

Berwick, Robert C. (1983), "Learning Word Meanings from Examples", *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83; Karlsruhe, West Germany)* (Los Altos, CA: William Kaufmann): 459–461.

Campbell, Alistair E. (1996), "Resolution of the Dialect Problem in Communication through Ontological Mediation", *Proceedings of the AAAI-96 Workshop on Detecting, Preventing, and Repairing Human-Machine Miscommunication (Portland, OR)*; ⟨http://www.cs.uwm.edu/∼mcroy/mnm.html⟩.

Campbell, Alistair E., & Shapiro, Stuart C. (1995), "Ontological Mediation: An Overview", *Proceedings of the IJCAI Workshop on Basic Ontological Issues for Knowledge Sharing (Montréal, Quebec, Canada)*.

The Compact Edition of the Oxford English Dictionary: Complete Text Reproduced Micrographically (New York: Oxford University Press, 1971).

Cravo, Maria R., & Martins, João P. (1993), "SNePSwD: A Newcomer to the SNePS Family", *Journal of Experimental and Theoretical Artificial Intelligence* 5: 135–148.

Duchan, Judith F.; Bruder, Gail A.; & Hewitt, Lynne E. (eds.) (1995), *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Ehrlich, Karen (1995), "Automatic Vocabulary Expansion through Narrative Context", *Technical Report 95-09* (Buffalo: SUNY Buffalo Department of Computer Science).

Ehrlich, Karen, & Rapaport, William J. (1997), "A Computational Theory of Vocabulary Expansion", *Proceedings of the 19th Annual Conference of the Cognitive Science Society (Stanford University)* (Mahwah, NJ: Lawrence Erlbaum Associates): 205–210.

Elshout-Mohr, Marianne, & van Daalen-Kapteijns, Maartje M. (1987), "Cognitive Processes in Learning Word Meanings", in Margaret G. McKeown & Mary E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 53–71.

Fodor, Jerry, & Lepore, Ernest (1992), *Holism: A Shopper's Guide* (Cambridge, MA: Basil Blackwell).

Gentner, Dedre (1981), "Some Interesting Differences between Nouns and Verbs", *Cognition and Brain Theory*, 4: 161–178.

Gleitman, Lila (1990), "The Structural Sources of Verb Meanings", *Language Acquisition* 1: 1–55.

Gleitman, Lila (1994), "A Picture is Worth a Thousand Words—But That's the Problem" (talk presented at the 1st International Summer Institute in Cognitive Science, SUNY Buffalo, July 1994); abstract in *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum Associates): 965.

Granger, Richard H. (1977), "Foul-Up: A Program that Figures Out Meanings of Words from Context", *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI-77, MIT)* (Los Altos, CA: William Kaufmann): 67–68.

Haas, Norman, & Hendrix, Gary (1983), "Learning by Being Told: Acquiring Knowledge for Information Management", in Riszard S. Michalski, Jaime G. Carbonell, & Tom M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach* (Palo Alto, CA: Tioga Press): 405–428.

Harman, Gilbert (1974), "Meaning and Semantics", in Milton K. Munitz & Peter K. Unger (eds.), *Semantics and Philosophy* (New York: New York University Press): 1–16.

Harman, Gilbert (1982), "Conceptual Role Semantics", *Notre Dame Journal of Formal Logic* 23: 242–256.

Hastings, Peter M. (1994), "Automatic Acquisition of Word Meanings from Natural Language Contexts", Ph.D. disseration (Ann Arbor: University of Michigan Department of Computer Science and Engineering).

Hastings, Peter M., & Lytinen, Steven L. (1994a), "The Ups and Downs of Lexical Acquisition", *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94, Seattle)* (Menlo Park, CA: AAAI Press/MIT Press): 754–759.

Hastings, Peter M., & Lytinen, Steven L. (1994b), "Objects, Actions, Nouns, and Verbs", *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum Associates): 397–402.

Higginbotham, James (1989), "Elucidations of Meaning", *Linguistics and Philosophy* 12: 465–517.

Hunt, Alan, & Koplas, Geoffrey D. (1997), "Definitional Vocabulary Acquisition Using Natural Language Processing and a Dynamic Lexicon", *SNeRG Technical Note 29* (Buffalo: SUNY Buffalo Department of Computer Science, SNePS Research Group).

Johnson-Laird, Philip N. (1987), "The Mental Representation of the Meanings of Words", in Alvin I. Goldman (ed.), *Readings in Philosophy and Cognitive Science* (Cambridge, MA: MIT Press, 1993): 561–583.

Kiersey, David M. (1982), "Word Learning with Hierarchy Guided Inference", *Proceedings of the National Conference on Artificial Intelligence (AAAI-82, Pittsburgh)* (Los Altos, CA: Morgan Kaufmann): 172–178.

Malory, Sir Thomas (1470), *Le Morte Darthur*, ed. by R. M. Lumiansky (New York: Collier Books, 1982).

Martins, João, & Cravo, Maria R. (1991), "How to Change Your Mind", *Noûs* 25: 537–551.

Martins, João, & Shapiro, Stuart C. (1988), "A Model for Belief Revision", *Artificial Intelligence* 35: 25–79.

Meinong, Alexius (1910), *Über Annahmen*, 3rd edition, based on the 2nd edition of 1910 (Leipzig: Verlag von Johann Ambrosius Barth, 1928).

Nutter, J. Terry (1983), "Default Reasoning Using Monotonic Logic: A Modest Proposal", *Proceedings of the National Conference on Artificial Intelligence (AAAI-83; Washington, DC)* (Los Altos, CA: Morgan Kaufmann): 297–300.

Quillian, M. Ross (1968), "Semantic Memory", in Marvin Minsky (ed.), *Semantic Information Processing* (Cambridge, MA: MIT Press): 227–270.

Quillian, M. Ross (1969), "The Teachable Language Comprehender: A Simulation Program and Theory of Language", *Communications of the ACM* 12: 459–476.

Rapaport, William J. (1981), "How to Make the World Fit Our Language: An Essay in Meinongian Semantics", *Grazer Philosophische Studien* 14: 1–21.

Rapaport, William J. (1988), "Syntactic Semantics: Foundations of Computational Natural-Language Understanding", in James H. Fetzer (ed.), *Aspects of Artificial Intelligence* (Dordrecht, Holland: Kluwer Academic Publishers): 81–131; reprinted in Eric Dietrich (ed.), *Thinking Computers and Virtual Persons: Essays on the Intentionality of Machines* (San Diego: Academic Press, 1994): 225–273.

Rapaport, William J. (1991), "Predication, Fiction, and Artificial Intelligence", *Topoi* 10: 79–111.

Rapaport, William J. (1995), "Understanding Understanding: Syntactic Semantics and Computational Cognition", in James E. Tomberlin (ed.), *AI, Connectionism, and Philosophical Psychology*, Philosophical Perspectives, Vol. 9 (Atascadero, CA: Ridgeview): 49–88; to be reprinted in Clark, Andy, & Toribio, Josefa (1998), *Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory*, Artificial Intelligence and Cognitive Science: Conceptual Issues, Vol. 4 (Hamden, CT: Garland).

Rapaport, William J.; Segal, Erwin M.; Shapiro, Stuart C.; Zubin, David A.; Bruder, Gail A.; Duchan, Judith F.; Almeida, Michael J.; Daniels, Joyce H.; Galbraith, Mary M.; Wiebe, Janyce M.; & Yuhan, Albert Hanyong (1989a), "Deictic Centers and the Cognitive Structure of Narrative Comprehension", *Technical Report 89-01* (Buffalo: SUNY Buffalo Department of Computer Science).

Rapaport, William J.; Segal, Erwin M.; Shapiro, Stuart C.; Zubin, David A.; Bruder, Gail A.; Duchan, Judith F.; & Mark, David M. (1989b), "Cognitive and Computer Systems for Understanding Narrative Text", *Technical Report 89-07* (Buffalo: SUNY Buffalo Department of Computer Science).

Rapaport, William J., & Shapiro, Stuart C. (1995), "Cognition and Fiction", in Judith F. Duchan, Gail A. Bruder, & Lynne E. Hewitt (eds.), *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates): 107–128; an abridged and slightly edited version appears as Rapaport, William J., & Shapiro, Stuart C. (forthcoming), "Cognition and Fiction: An Introduction", in Ashwin Ram & Kenneth Moorman (eds.), *Understanding Language Understanding: Computational Models of Reading* (Cambridge, MA: MIT Press).

Rapaport, William J.; Shapiro, Stuart C.; & Wiebe, Janyce M. (1997), "Quasi-Indexicals and Knowledge Reports", *Cognitive Science* 21: 63–107.

Rosch, Eleanor (1978), "Principles of Categorization", in Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and Categorization* (Hillsdale, NJ: Lawrence Erlbaum Associates): 27–48.

Schank, Roger C., & Abelson, Robert P. (1977), *Scripts, Plans, Goals and Understanding* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Searle, John R. (1980), "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3: 417–457.

Segal, Erwin M. (1995), "A Cognitive-Phenomenological Theory of Fictional Narrative", in Judith F. Duchan, Gail A. Bruder, & Lynne E. Hewitt (eds.), *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates): 61–78.

Sellars, Wilfrid (1955), "Some Reflections on Language Games", in *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963): 321–358.

Shapiro, Stuart C. (1979), "The SNePS Semantic Network Processing System", in Nicholas Findler (ed.), *Associative Networks: Representation and Use of Knowledge by Computers* (New York: Academic Press): 179–203.

Shapiro, Stuart C. (1982), "Generalized Augmented Transition Network Grammars for Generation from Semantic Networks", *American Journal of Computational Linguistics* 8: 12–25.

Shapiro, Stuart C. (1989), "The CASSIE Projects: An Approach to Natural Language Competence", *Proceedings of the 4th Portugese Conference on Artificial Intelligence (Lisbon)* (Berlin: Springer-Verlag): 362–380.

Shapiro, Stuart C., & Rapaport, William J. (1987), "SNePS Considered as a Fully Intensional Propositional Semantic Network", in Nick Cercone & Gordon McCalla (eds.), *The Knowledge Frontier: Essays in the Representation of Knowledge* (New York: Springer-Verlag): 262–315; shorter version appeared in *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86, Philadelphia)* (Los Altos, CA: Morgan Kaufmann): 278–283; revised version of the shorter version appears as "A Fully Intensional Propositional Semantic Network", in Leslie Burkholder (ed.), *Philosophy and the Computer* (Boulder, CO: Westview Press, 1992): 75–91.

Shapiro, Stuart C., & Rapaport, William J. (1991), "Models and Minds: Knowledge Representation for Natural-Language Competence", in Robert Cummins & John Pollock (eds.), *Philosophy and AI: Essays at the Interface* (Cambridge, MA: MIT Press): 215–259.

Shapiro, Stuart C., & Rapaport, William J. (1992), "The SNePS Family", *Computers and Mathematics with Applications* 23: 243–275; reprinted in Fritz Lehmann (ed.), *Semantic Networks in Artificial Intelligence* (Oxford: Pergamon Press, 1992): 243–275.

Shapiro, Stuart C., & Rapaport, William J. (1995), "An Introduction to a Computational Reader of Narrative", in Judith F. Duchan, Gail A. Bruder, & Lynne E. Hewitt (eds.), *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates): 79–105.

Siskind, Jeffrey Mark (1996), "A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings", *Cognition* 61: 39–91.

Srihari, Rohini K. (1991), "PICTION: A System that Uses Captions to Label Human Faces in Newspaper Photographs", *Proceedings of the 9th National Conference on Artificial Intelligence (AAAI-91, Anaheim)* (Menlo Park, CA: AAAI Press/MIT Press): 80–85.

Srihari, Rohini K., & Rapaport, William J. (1989), "Extracting Visual Information From Text: Using Captions to Label Human Faces in Newspaper Photographs", *Proceedings of the 11th Annual Conference of the Cognitive Science Society (Ann Arbor, MI)* (Hillsdale, NJ: Lawrence Erlbaum Associates): 364–371.

Sternberg, Robert J. (1987), "Most Vocabulary is Learned from Context", in Margaret G. McKeown & Mary E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 89–105.

Winston, Patrick Henry (1975), "Learning Structural Descriptions from Examples", in Patrick Henry Winston (ed.), *The Psychology of Computer Vision* (New York: McGraw-Hill): 157–209; reprinted in Ronald J. Brachman & Hector J. Levesque (eds.), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kaufmann, 1985): 141–168.

Zadrozny, Wlodek, & Jensen, Karen (1991), "Semantics of Paragraphs", *Computational Linguistics* 17: 171–209.

Zernik, Uri, & Dyer, Michael G. (1987), "The Self-Extending Phrasal Lexicon", *Computational Linguistics* 13: 308–327.