A COMPUTATIONAL THEORY OF VOCABULARY EXPANSION
Project Proposal

Karen Ehrlich and William J. Rapaport

Department of Computer Science and Center for Cognitive Science
State University of New York at Buffalo, Buffalo, NY 14260

{ehrlich|rapaport}@cs.buffalo.edu

Abstract

This project concerns the development and implementation of a computational theory of how human readers and other natural-language-understanding systems can automatically expand their vocabulary by determining the meaning of a word from context. The word might be unknown to the reader, familiar but misunderstood, or familiar but being used in a new sense. 'Context' includes the prior and immediately surrounding text, grammatical information, and the reader's background knowledge, but no access to a dictionary or other external source of information (including a human). The fundamental thesis is that the meaning of such a word (1) *can* be determined from context, (2) can be *revised* and refined upon further encounters with the word, (3) "*converges*" to a dictionary-like definition if enough context has been provided and there have been enough exposures to the word, and (4) eventually "*settles down*" to a "steady state", which, however, is always subject to revision upon further encounters with the word. The system is being implemented in the SNePS-2.1 knowledge-representation and reasoning system, which provides a software laboratory for testing and experimenting with the theory. This research is a component of an interdisciplinary, cognitive-science project to develop a computational cognitive model of a reader of narrative text.

0 SUMMARY. This project concerns the development and implementation of a computational theory of how human readers and other natural-language-understanding systems can automatically increase their lexicon (i.e., expand their vocabulary) by determining the meaning of a word from context. The word might be unknown to the reader, familiar but misunderstood, or familiar but being used in a new sense. 'Context' includes the prior and immediately surrounding text, grammatical information, and the reader's background knowledge, but no access to a dictionary or other external source of information (including a human).

The fundamental thesis is that the meaning of such a word (1) *can* be determined from context, (2) can be *revised* and refined upon further encounters with the word, (3) "*converges*" to a dictionary-like definition if enough context has been provided and there have been enough exposures to the word, and (4) eventually "*settles down*" to a "steady state", which, however, is always subject to revision upon further encounters with the word. Each encounter with the word yields a definition—a hypothesis about meaning. Each subsequent encounter provides an opportunity to revise this hypothesis in the light of new evidence. The revision is *unsupervised*: There is no (human) "trainer" and no "error-correction" techniques. Finally, no *domain-specific* antecedent background information is required for the development and revision of the hypothesized definition (with the exception of the word's lexical category (part of speech)).

The system is being implemented in the SNePS-2.1 knowledge-representation and reasoning system. SNePS's inference package allows rules for both deductive and default reasoning. In the presence of a contradiction, a belief-revision package allows the user to remove from the context in which the contradiction arose one or more of the propositions from which the contradiction was derived. Once the offending premise is no longer asserted, the conclusions that depended on it also cease to be asserted in that context. This mechanism is used to revise definitions that are inconsistent with a word's current use. SNePS also has an English lexicon, morphological analyzer/synthesizer, and a generalized ATN parser-generator that translates the input English directly into a propositional semantic network.

The vocabulary-expansion system, called 'Cassie', consists of SNePS-2.1 and a knowledge base of background information. Cassie's input consists, in part, of information from the text being read, which is input directly in the knowledge-representation formalism. A major part of the proposed research is updating and further developing the grammar in order to automate the transduction of sentences from the text into information in the knowledge base. The sentences themselves will still be entered by hand, although experiments are planned using on-line text corpora. Cassie's other input is questions asked about the material being read. In particular, we can ask, "What does ⟨word⟩ mean?" This triggers a deductive search of the knowledge base, consisting of background information plus information from the story, all marked with its "degree" of immunity from revision. Output consists of a report of Cassie's current definition of the word, or answers to other queries.

This system, which provides a software laboratory for testing and experimenting with the theory, is a component of an interdisciplinary, cognitive-science project to develop a computational cognitive model of a reader of narrative text. The proposed research will be an important contribution to this and similar projects, since to fully model 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 dictionary. Moreover, a complete lexicon could not be manually encoded, nor could it contain neologisms or new meanings given to old words in new contexts. Text-understanding, message-processing, or information-extraction systems need to be robust: They should not break down just because they have encountered an unknown expression. This is especially the case for systems that use unconstrained input text and must operate independently of human intervention. One application is to "intelligent agents", which ought to be able to figure out a human user's instructions without necessarily stopping to ask what each new word means. Other possible applications include language-acquisition studies and computational lexicography. Issues to be investigated include: elaborating the algorithms for defining verbs, modifiers, and proper names; dealing with synonyms and ill-formed input; using etymological and morphological clues for definitions; inducing generalizations from instances; and determining which parts of a hypothesis to revise.

1 INTRODUCTION. We propose to continue our development of a computational theory of how readers (or natural-language-understanding (NLU) systems) can automatically increase their lexicon (i.e., expand their vocabulary) by determining the meaning of a word from context [20]. The word might be unknown to the reader, familiar but misunderstood, or familiar but being used in a new sense. By 'context', we include the prior and immediately surrounding text, grammatical information, and the reader's background knowledge, but no access to a dictionary or other external source of information (including a human).

Our fundamental thesis is that the meaning of such a word (1) *can* be determined from context, (2) can be *revised* and refined upon further encounters with the word, (3) "*converges*" to a dictionary-like definition if enough context has been provided and there have been enough exposures to the word, and (4) eventually "*settles down*" to a "steady state", which, however, is always subject to revision upon further encounters with the word. Each encounter with the word yields a definition—a hypothesis about meaning. Each subsequent encounter provides an opportunity to revise this hypothesis in the light of new evidence. The revision is *unsupervised*: There is no (human) "trainer" and no "error-correction" techniques. The hypothesized definitions are *not* guaranteed to converge to a (or the) "correct" meaning of the word (if such exists). However, they *do* converge on a meaning that is stable with respect to further encounters. Finally, no *domain-specific* antecedent background information is required for the development and revision of the hypothesized definition (with the exception of the word's lexical category (part of speech)).

Evidence for our thesis can be seen in the psychological literature and in (informal) protocols that we have taken from subjects who were asked to reason out loud about their definition-forming and revision procedures when shown passages containing unknown words. These same passages have served as input to a computational system that develops and revises definitions in ways similar to the human subjects.

This system serves two purposes: (1) It provides a software laboratory for testing and experimenting with the detailed implementation of our theory. (2) It is a component of a larger project to develop a computational cognitive model of a reader of narrative text, summarized in [19] (see also: [1] [5], [7], [17], [19], [22], [35], [45], [48], [50]–[53], [60]–[62], [70], [73]–[77], [85], [90]–[92], [94] [96]).

2 SIGNIFICANCE. The proposed research will be an important contribution to this and similar projects, since to fully model a reader, it is important to model the ability to learn from reading ([62], [64]), in particular, to expand one's vocabulary in a natural way while reading, without having to stop to ask someone or to consult a dictionary.

This research is also of independent significance. It is part of the task of lexical processing of text, dealing, in particular, with the need to process new or unknown words. As pointed out in [98], no assumption of a "fixed complete lexicon" can be made: It could not be manually encoded, nor could it contain neologisms or new meanings given to old words in new contexts. As an issue in computational linguistics, we want NLU systems (a.k.a. "text-understanding", "message-processing", or "information-extraction" systems) to be (as) robust (as humans): They should not break down just because they have encountered an unknown expression. This is especially the case for systems that use unconstrained input text and must operate independently of human intervention. E.g., "intelligent agents" ought to be able to figure out a human user's instructions without necessarily stopping to ask what each new word means. Similarly, a system designed to locate "interesting" news items from an online information server should not be limited to keyword searches; e.g., if the user is interested in news items about dogs, 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 a kind of dog.

Two key features of our system mesh nicely with these desiderata; they can be summarized as the advantages of learning over being told: (1) Being told requires human intervention, whose availability cannot be guaranteed. *Our system operates independently of a human teacher or trainer* (with one exception that we propose to overcome). (2) One can't predict all information that might be needed to understand unconstrained, domain-independent text; hence, the system must be able to learn on its own. *Our system does not constrain the subject matter (the "domain") of the text.* Although we are primarily concerned with narrative text, our techniques are perfectly general; we propose to extend them to other kinds of texts and knowledge-sharing applications. Assuming the availability of an appropriate grammar, we are developing (and propose to elaborate on) algorithms for producing definitions independent of domain. However, the definitions *are* dependent on the system's background knowledge: The more background knowledge the system has, the better the definitions will be, and the more quickly they will "converge". We are *not* proposing a system that develops "correct" definitions (see §3); rather, we propose a system that develops dictionary-like definitions that enable the reader to continue with the task of understanding the text.

Other possible applications of our system include its use in language-acquisition studies in cognitive science, since our implemented model will be a good "laboratory" for comparing vocabulary-learning behaviors of humans and computers. We plan also to investigate its applicability to computational lexicography: i.e., the automatic construction of dictionary entries given samples of use-in-context of words to be defined.

3 THEORETICAL BACKGROUND. Our implemented theory at least partially tests the thesis that symbol manipulation (syntax) suffices for NLU [60], [63]. Humans understand one another by interpreting the symbols they read or hear. This interpretation is a mapping from the speaker's (or writer's) syntax to the hearer's (or reader's) concepts (semantics). We take the meaning of a word (as understood by a cognitive agent) to be the position of that word in a highly interconnected network of words, propositions, and other concepts. I.e., a word's meaning is its (syntactic) relation to other words, and words of similar meaning will have similar connections [55], [56]. However, even words whose meanings are very similar to one another do not have identical connections to other terms and are, therefore, distinguishable from one another in the context of the agent's experience. We thus adopt Quine's view that the beliefs held by a cognitive agent form an interconnected web, where a change or addition to some portion of the web can affect other portions that are linked to it [57]. Such an agent's understanding of natural-language (NL) input will, therefore, be a part of such a web or semantic network composed of internal (mental) objects. If we take these mental objects to be symbols, then the interpretation of linguistic input is a syntactic operation, and formal symbol manipulation is sufficient for attaching meanings to words [60], [63].

Objects about which we can think, speak, or write need not exist. On a purely referential, or extensional, semantics, words for such non-existent objects would be meaningless. However, in a syntactically based semantics, the linguistic contexts in which words for such objects are found can provide meanings for them. The meaning of an expression (for an agent, in a context) can be viewed as a mental object made of the internal representation of fragments of linguistic input previously encountered by the agent. E.g., the meaning of 'bachelor' for agent A in context C might be ⟨_s are unmarried, John is a _, that guy is a _, no women are _s, _s are men, etc.⟩. As different uses of a word are heard or read, new contexts come into being. Since we take meaning to be context, an agent's understanding of a word's meaning is thus revised by successive encounters with it [58].

In this (idiolectic) sense, the meaning of a word for a cognitive agent is determined by idiosyncratic experience with it. In another sense, the meaning of a word is its dictionary definition, which usually contains less information than the idiolectic meaning. Taken to its fullest extent, the contextual meaning described above includes a word's relation to every concept in the agent's mind. The meaning of 'bachelor' would involve all the concepts connected to 'unmarried', 'John', 'that guy', 'women', 'men', and all the concepts connected to those concepts, etc., throughout the entire web of the agent's thoughts. Thus, unless the agent has two completely unrelated networks of ideas, the extreme interpretation of "meaning as context" defines every word in terms of every other word an agent knows. This is too circular and too unwieldy to be of much use. We need to limit the connections used to provide the definition.

In a related project [36], Hill proposed a SCOPE function that selects a subnetwork containing only those concepts reachable from a particular concept by a path of connections that is no longer than a given length. Consider a case where agent A has direct links from 'bachelor' to 'John', 'John' to 'tall', and 'tall' to 'big'. Suppose there is no shorter path from 'bachelor' to 'big' than that which runs through 'John' and 'tall'. Then, if b represents A's concept of 'bachelor', the function SCOPE(b, 2) would select from A's mental network a subnetwork containing A's concepts of 'John' and 'tall', but not 'big'. SCOPE(b, 3) would include 'big', but not other concepts reachable from it, unless they had other, shorter paths to 'bachelor'.

Rather than selecting a subnetwork of a particular *scope* in order to limit the connections used to provide a definition, we select for particular *kinds* of information. Not all concepts within a given subnetwork need be equally salient to a conventional, dictionary-style definition of a word. Agent A may have a direct mental connection from 'bachelor' to 'John', but agent B may never have heard of him, yet A and B may be able to agree on a definition of 'bachelor'. In the attempt to understand and be understood, people abstract certain conventional information about words and accept this information as a definition. When a new word is encountered, people begin to hypothesize a definition.

There are two approaches to lexicography: On the prescriptive approach, the definitions given are the "correct" definitions, and any other use is, by definition, incorrect. The other approach is descriptive: A word means what it is used to mean by those who use it. E.g., the prescriptive definition of the phrase '*à la mode*' is "in the current fashion", but a descriptive definition of the phrase as observed in American contexts would be something like "with ice cream". Our system produces descriptive definitions (cf. §8.7).

4 IMPLEMENTATION.

Our system is being implemented in the SNePS-2.1 semantic-network knowledge-representation and reasoning (KRR) system developed by S.C. Shapiro and the SNePS Research Group [71], [76], [78], [80]. SNePS has been and is being used for a number of research projects in NLU [1]–[2], [48]–[53], [59]–[60], [62]–[63], [65], [72]–[73], [76]–[77], [79], [89]–[94], [96]–[97].

SNePS is an appropriate KRR system for our approach to lexical semantics. Each node in a SNePS network represents a concept or mental object (possibly built of other concepts), with labeled arcs linking the concepts. All information, including propositions, is represented by nodes, and propositions about propositions can be represented without limit. Arcs merely form the underlying syntactic structure of SNePS. This is embodied in the restriction that one cannot add an arc between two existing nodes. That would be tantamount to telling SNePS a proposition that is not represented by a node. Paths of arcs can be defined, allowing for path-based inference, including property inheritance within generalization hierarchies. There is a 1–1 correspondence between nodes and represented concepts. This uniqueness principle guarantees that nodes will be shared whenever possible and that nodes represent intensional objects such as concepts, propositions, properties, algorithms, and objects of thought including fictional entities (e.g., Sherlock Holmes), non-existents (e.g., the golden mountain), and impossible objects (e.g., the round square) [76], [77].

SNePS's inference package allows one to write rules for ordinary deductive reasoning as well as for default reasoning, which allows the system to infer "probable" conclusions in the absence of specific information to the contrary. When certain combinations of asserted propositions lead to a contradiction, the SNeBR belief-revision package allows the user to remove from the context in which the contradiction arose one or more of the propositions from which the contradiction was derived [47]. Once the offending premise is no longer asserted, the conclusions that depended on it also cease to be asserted in that context. We use this mechanism to revise those beliefs about the meanings of words that turn out, upon further encounters with the words, to be inconsistent with their use.

We have developed algorithms for *partially* automating the identification and removal or modification of the offending premise, based on SNePSwD, a default belief-revision system that enables automatic revision [16], [46]. We propose to explore techniques for *fully* automating it.

SNePS also has an English lexicon, morphological analyzer/synthesizer, and a generalized ATN parser-generator that, rather than building an intermediate parse tree, translates the input English directly into a propositional semantic network ([72], [73]; see [60], [62], [79] for detailed examples).

5 CURRENT STATUS. "Cassie", our vocabulary-expansion system, consists of SNePS-2.1 (including SNeBR and the ATN parser-generator), SNePSwD, and a knowledge base (KB) of background information. Currently, the KB is hand-coded, because it represents Cassie's antecedent knowledge; *how* she acquired this knowledge is irrelevant. We begin with what some might call a "toy" KB, but each of our tests so far has included all previous information, so the KB grows as we test more words. This raises the problem of controlling and localizing the reasoning process. Although we do not have a panacea for this, some results have been obtained in a related project on SNePS reasoning [11]; see also §§8.2, 8.3.

Cassie's input consists, in part, of information from the text being read. Currently, this, too, is input directly in the KRR formalism. However, a major part of our proposed research is the updating and further development of our grammar in order to automate the transduction of sentences from the text into information in the KB. The sentences themselves will still be entered by hand, although we plan to experiment with using an on-line text corpus that could be interfaced with SNePS, such as the tagged *Wall Street Journal* (WSJ) corpus, the (tagged) Brown Corpus, or other corpora available from sources such as the Linguistic Data Consortium. Cassie's other input is questions asked about the material being read. In particular, we can ask, "What does ⟨word⟩ mean?" This triggers a deductive search of the KB, consisting of background information plus information from the story, all marked with its "degree" of immunity from revision. Output consists of a report of Cassie's current definition of the word, or answers to other queries.

5.1 Algorithms. At present, we have developed algorithms for hypothesizing and revising meanings for nouns and verbs that are unknown, mistaken, or being used in a new way. Cassie was provided with background information for understanding the King Arthur stories in the *Morte Darthur* [43]. E.g., when presented with a sequence of passages involving the hitherto unknown noun 'brachet', Cassie was able to develop a theory that a brachet was a dog whose function is to hunt and that can bay and bite. (*Webster's Second* [88] defines it as "a hound that hunts by the scent".) However, based on the first context in which the term appeared (viz., "Right so as they sat, 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 with a great cry."), the initial hypothesis was merely that a brachet was a physical object that may be white. Each time the term appeared, Cassie was asked to define it. To do so, she deductively searched her background KB, 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, the less salient "other properties" of particular brachets are not reported, although we do report any 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 psycholinguistic studies of the sort of vocabulary expansion we are modeling [21], [39], [86]. In the absence of some or all 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.

In another test, Cassie was told that 'to smite' meant "to kill by hitting hard" (a mistaken belief actually held by the PI before reading [43]). Passages in which various characters were smitten but then continued to act triggered SNeBR, which asks the user (i.e., the reader) which of several possible "culprit" propositions in the KB to remove in order to block inconsistencies. Ideally, the reader then decides which belief to revise. Although the decision about which proposition (representing an incorrect definition) to withdraw and which *new* proposition (representing a revised definition) to add has been partially automated (using SNePSwD), this remains the one area still occasionally requiring human intervention. Automating this will be a major focus of our research.

A third case is exemplified by 'to dress', which Cassie antecedently understood to mean "to put clothes on (something)". This is a well-entrenched meaning, which should *not* be rejected. However, upon reading that King Arthur "dressed" his sword, SNeBR detects an inconsistency. Rather than *rejecting* the prior definition, we *add* to it. In this case, Cassie decides that to dress is *either* to put clothes on *or* to prepare for battle. We plan to investigate getting Cassie to *generalize* in situations like this: Since getting dressed is also a form of preparation (as is the use of salad "dressing"), Cassie should be able to *induce* a more general meaning (while maintaining such everyday meanings as putting clothes on).

Applying the principle that the meaning of a term is its location in the network (here, a network of background information and story information), our algorithms for defining terms deductively search the network for information appropriate to a dictionary-like definition. The following sections sketch our algorithms for defining nouns and verbs and for revision. We assume that our grammar has been able to identify the unknown word as a noun or a verb.

5.2 Noun-Defining Algorithm. Let N be the unknown noun.

```
PROCEDURE List1 ::= list (1) structure of Ns, (2) functions of Ns,
                         (3) stative properties of Ns only if there are general rules about
                             them.

PROCEDURE 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'.

PROCEDURE List3 ::= BEGIN List2;
                          IF the system finds structural or functional information about Ns,
                             THEN List1 END.

BEGIN
  IF N represents a basic-level category, THEN 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 List3
  ELSIF N represents a subclass of physical object, THEN
    BEGIN List2;
          IF system finds structural or functional information about Ns, THEN 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: any 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.
```

In all cases, we try to find functional information, and in all cases except that of abstract nouns, we look for structural information. If such information can be found, we look for stative properties only as general rules. We do not seek stative properties of individuals once we have data about structure or function unless we are lacking any class inclusions, because such descriptions are not usually salient to a dictionary-style definition. If we have no class inclusions, and are working solely with individuals, we report stative descriptive properties even if we find some structural or functional information about the individuals: When we have no class inclusions, we need any information we can get. However, we do not bother trying to find synonyms, because we cannot make any reasonable guesses at synonyms without class inclusions. We look for actions attributed to Ns or, failing that, to individuals of class N in all cases except where N represents a variety of an inanimate basic-level category. (If we know that we're talking about a type of chair, there's no reason to look for actions, but if we know only that we're talking about a physical or abstract object, we don't know enough to rule out the possibility that it can act.) If our class inclusions are either limited to vague categories such as physical or abstract object, or lacking altogether, *and* if we can find no information about structure, function, or actions (i.e., if the only slots in our standard framework that may be filled are stative properties and possible ownership of individuals), then we look for occurrences of the target word as the object of an act performed by an agent.

Deciding what information to look for is largely controlled by what we know about class inclusion, especially whether the target noun represents a subordinate, superordinate, or basic-level category [66]. We *infer* this information to make certain that any implicit class inclusions are noticed by the system. Once the deduction is made, however, we *search* the network to abstract the classes. All other definitional information is also abstracted by finding specific paths in the network.

5.3 Verb-Defining Algorithm. Let V be the unknown verb.

```
BEGIN
  report on cause and effect; categorize the subject;
  IF V is used with an indirect object, THEN categorize objects and 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.
```

To define V, we currently report its predicate structure, a categorization of its arguments, and any causal or enablement information we can find. To categorize an argument of V, we look for, in order of preference, its membership in a basic-level category, membership in a subclass of animal, or membership in some other known category.

5.4 Revision. When humans encounter a discrepancy between the way a word is used and their previous understanding of it, they must either assume that the word is used incorrectly or decide that their previous understanding requires revision. When Cassie encounters a contradiction derived from combining story information with background knowledge, she must decide which of the premises leading to the contradiction should be revised. To facilitate this, we tag each assertion in the KB and the story with a "knowledge category" (`kn_cat`). (Assertions having no `kn_cat` are beliefs Cassie has derived.) These are ordered in a hierarchy of certainty of belief, so that the system can restrict the field from which she chooses a belief for revision to premises believed with the least certainty. The hierarchy of `kn_cats`, from greatest certainty of belief to least, is:

1. `kn_cat` *intrinsic*: Facts about language, including simple assertions and rules; very basic or fundamental background knowledge. Found in the KB, not (usually) in stories. E.g., the temporal relation "before" is transitive; containment of an item in a class implies containment in superclasses; encountering the usage V(agent, object, indobj) implies that V 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 it); the SNePS representation that would be built on parsing a sentence in the story. E.g., "Sir Gryflette left his house and rode to town" contains the following story facts: Someone is named Sir Gryflette. That someone left his house. That someone rode to town.

3. `kn_cat` *life*: Background knowledge expressed as simple assertions without variables or inference. E.g., taxonomies (e.g., dogs are a subclass of animals), assertions about individuals (e.g., Merlin is a wizard).

4. `kn_cat` *story-comp*: Information not directly present in the story, but inferred by the reader to make sense of it. Such "story completion" [69] 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 Cassie. We include it for cases where a gap in her KB might leave her unable to infer some fact necessary to understanding the story. Using the example from the `kn_cat` *story*, 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) reflecting common, everyday knowledge. E.g., 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) reflecting specialized, non-everyday information. E.g., 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. E.g., 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. Cassie attaches this tag when she revises a rule that was tagged as a *life-rule.2*. It is a temporary classification while Cassie looks for confirmation of her revision. Once she settles on a particular revision, the revised rule is tagged as a *life-rule.2*.

In case of contradiction, Cassie selects, from among the conflicting propositions, a proposition of greatest uncertainty as a candidate for revision. If only one belief has the highest level of uncertainty in the conflict set, it will be revised. If several alternatives exist with the same (highest present) `kn_cat`, Cassie looks for a verb in the antecedent (humans more readily revise beliefs about verbs than about nouns [24]). If this is still insufficient to yield a single culprit, then, in the current implementation, a human "oracle" chooses one (although this hasn't been needed in our tests to date). Ideally, Cassie would use discourse information to make the decision between possible culprits at the same level of certainty. E.g., in the case of dressing a sword before fighting, the rule about what it means to dress something might be selected for revision because it is unrelated to the topic of fighting, whereas swords are closely associated with the topic. We will explore refinements of this hierarchy as part of the proposed research.

At present, all the `kn_cat`s (except for *questionable*) are assigned by a human when the proposition is input. The assignment of the `kn_cat` *story* could be handled automatically: Cassie would simply include it as a part of each proposition built from the parse of a sentence in a story [62], [64]. Since *story-comp* is only a stop-gap measure, we need not worry about how Cassie might assign it: Either she wouldn't make any such assignment, or she would tag all derived propositions as being derived as is already done by SNeBR. How might Cassie categorize the non-derived assertions in her KB? 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. We plan to investigate how such intuitions can be formalized and automated.

Finally, a set of rules for replacing discarded definitions with revised definitions is being developed. Here are two samples of such rules:

> If the culprit has `kn_cat` *life-rule.2*, and if there are multiple consequents, $cq_1\& \ldots \&cq_n$, and if cq_n says that cq_i is the cause of cq_j, and if cq_j is contradicted, then substitute two rules. The first says that the original antecedent implies $cq_1\& \ldots \&\mathrm{Possibly}(cq_j)\& \ldots \&cq_{n-1}$ and has `kn_cat` *questionable*. The second conjoins the original antecedent with cq_j to form a new antecedent, has as its only consequent cq_n, and has `kn_cat` *life-rule.2*.

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 (`kn_cat` *life-rule.2*). On first encountering a smitee who survives, substitute the pair of rules: (1) If x smites y, then x hits y & possibly y is dead (`kn_cat` *questionable*); (2) If x smites y and y is dead, then x hitting y causes y to be dead (`kn_cat` *life-rule.2*). 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.

> If the culprit has `kn_cat` *life-rule.1* and has a single consequent (that is not a disjunction), then substitute a version of the culprit rule with a consequent that is an exclusive disjunction of the old consequent with the place-holder SOMETHING.

E.g., suppose we had a rule: If x dresses y, then x puts clothes on y (`kn_cat` *life-rule.1*). On encountering a contradiction, substitute the rule: If x dresses y, then x puts clothes on y xor SOMETHING (`kn_cat` *life-rule.1*). This is used when the definition is well-entrenched, and we are adding a new meaning.

6 AN EXAMPLE. Here, we sketch Cassie's handling of 'smite', with this background information in the KB: *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 KB.)

Cassie is then given a sequence of passages containing 'smite' ([43]: 13ff) interspersed with questions and requests for definitions (**P**n = passage #n; **D**n = the definition created after **P**n; **Q**n = a question asked after **P**n; **R**n = the answer.)

P1: King Arthur turned himself and his horse. He smote before and behind. His horse was slain. King Lot smote down King Arthur.
D1: 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.
Q1: What properties does King Arthur have?
R1: King Arthur is dead.
P2: King Arthur's knights rescued him. They sat him on a horse. He drew Excalibur.
Q2: When did King Arthur draw?

The inference required to reply to **Q2** 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 (`kn_cat` *questionable*), and that if smiting results in a death, then the hitting is the cause of death. (These rules are *not* built in, ready to be called when needed. They are *inferred* by the sorts of revision rules discussed in §5.4.)

D2: 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.
P3: 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 that they were hurt and therefore could not have died earlier at the time that 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:

D3: 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. Not only has the definition stabilized ("converged"), but it has done so in a manner similar to our human protocols.

7 RELATED WORK. We have endeavored to incorporate, or be consistent with, research in psycholinguistics on language acquisition, notably the work already cited in §§5.1, 5.4. We plan to elaborate our theory of verb acquisition in the light of research on verb meanings [25], [26].

In computational linguistics, there have been many language acquisition and definition projects. In addition to those already cited, there are: [3], [4], [9], [10], [27]–[33], [40], [54], [81]–[83]. There are, of course, many more. In this section, we can only briefly discuss a few of these.

7.1 Zernik & Dyer [98] compiles definitions of words and phrases, especially figurative phrases, from conversation into a hierarchical lexicon. Figurative phrases are linguistic patterns whose meaning cannot be understood by composition of the meanings of their component words (e.g., "to put one's foot down"). Their system has a pattern constructor that analyzes parsing failures and modifies its patterns accordingly, and a concept constructor that selects from a set of strategies 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. It then suggests various interpretations more consistent with the goals of persons in the story until the user confirms that a correct interpretation has been reached.

Since we focus on literary, rather than conversational, discourse, Cassie is not informed by a human user when she misunderstands. This does not mean that her errors will always remain undetected. As long as the misunderstanding is compatible with further encounters with the word and with Cassie's general knowledge, there is no reason for Cassie to revise her understanding. However, if further reading leads to the conclusion that a previous definition was in error, Cassie revises her understanding without explicit instruction.

We have not attempted to deal with figurative phrases or idioms. Cassie treats all words literally. Some multi-word lexical entries exist, but they are treated as words, rather than as phrases, and their meanings are not derived from (nor even necessarily related to) the meanings of their component words. If the literal interpretation of a word that was used figuratively should cause a contradiction with Cassie's general knowledge, she may revise her prior understanding of the word by creating a secondary meaning. Such a secondary meaning might correspond to the figurative sense of the word, but Cassie would not know that it was figurative, nor would she be able to construct a new multi-word lexical entry from input of a series of single words. We propose to explore these possibilities further.

7.2 Pustejovksy [54] has discussed defining unknown verbs, using an "Extended Aspect Calculus", in terms of their associated case-roles or arguments, using certain semantic and syntactic markers, including a notion of canonical English word order. He also uses simulated visual information in conjunction with his Calculus, so that his system "sees" the action of a sentence while "hearing" it. This was an attempt to model a child gaining an understanding of a verb by observing an "extensional definition". His focus in this research was on how a child might identify particular thematic roles with specific grammatical functions in a sentence. Thematic roles are not assigned randomly to verbs, but exist in a hierarchy. A verb will not, for example, have a *Goal* unless it first has a *Theme*. In addition to the dependencies among roles, meaning can be marked as positive or negative with respect to motion, abstraction (tangibility vs. intangibility), directness, completeness, and animacy. The theory is that the thematic roles and the markedness can be perceptually observed, and used to understand a sentence by mapping the roles to the parts of the sentence.

Thus, if Pustejovsky's system, TULLY, is informed of an animate agent with direct motion and an animate object negative as to motion (at beginning, middle, and end times) while hearing the sentence: "Mary hit the cat", TULLY will map Mary to the role of agent and the cat to the role of patient, and so understand the sentence.

Because we focus on how a linguistically competent agent may learn new words, rather than on how children initially acquire language, we are not concerned with how to map thematic roles to words in a sentence. We assume Cassie has a sufficiently sophisticated knowledge of grammar that such assignments are made automatically while reading. We do, however, use information about the argument structure of a verb, and about what types of nouns can serve as arguments to a particular verb, in synthesizing its definition. We do not use any form of extensional definition, since one almost never has such information available when one encounters a new word in narrative. Cassie must produce her definitions of verbs, as of other words, strictly from NL context and stored knowledge.

7.3 Hastings [30]–[32] 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 in that the goal is to allow the system to read and acquire word meanings "on its own", 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 KR.

For each domain, the initial KB consisted of a taxonomy of relevant objects and actions. Camille attempts to place the unknown word appropriately in the domain hierarchy. In this respect, Camille can be viewed as an implementation of the theory of [40]. To this basic system, Hastings has added: a mutual exclusivity constraint (to avoid mapping too many words to the same concept); a script applier that allows Camille to match the unknown word with a known word that has filled the same slot in a particular script, or, in the case of a verb, with a known word whose arguments (slot-fillers) match those of the target word; and an ability to recognize the existence of multiple senses for a word (so that the same word may fit more than one place in the domain hierarchy). 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. The focus of the work 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.

7.4 Siskind [82], [83] is concerned with learning the meanings of words, as we are, but his focus is on first-language acquisition during childhood, whereas ours is on relatively mature cognitive agents who already know a large part of their language and are (merely) expanding their vocabulary. Nonetheless, some of our techniques are applicable to the sorts of cases Siskind considers; we plan to investigate this further. There are, however, important differences in our approaches. 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. He has implemented algorithms that 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 (a sentence from the text being read) and a mental representation that the utterance means (the SNePS network built as a result of reading the sentence), Cassie does not use visual perception to produce the mental representation. In most cases of reading, any mental representation (including any mental imagery) would be produced by the text, so visual perception of a real-world situation simply does not arise. There is, however, an obvious exception: illustrated texts. Although Cassie currently does not make use of any illustrations, she could (in principle): First, assuming appropriate techniques for transducing illustrations to (mental) representations in the KB, such representations could then be used to help determine meanings, since they would be part of the KB. This could be done in one of at least two ways: (1) by using *propositional* representations of the illustrations, in which case our algorithms would work more or less as they do now, or (2) by having explicit reference to *imagistic* representations, including these as part of the meaning [38]. Second, we have already experimented with techniques for merging illustrations and text, so we already have mechanisms in place for carrying this out [84], [85].

Another difference between our systems is that Siskind's 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 instead seeks to find or infer definition-style relations between the unknown word and the rest of the KB. Moreover, as explained in §§3, 8.7, 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, our system uses *syntactic* bootstrapping (using syntax to aid in learning semantics [25], [26]), which seems more reasonable for our situation.

8 NEXT STEPS.

8.1 Developing the Grammar. One significant improvement that we must undertake is the development of a generalized ATN grammar that can parse simplified forms of the sentences in [43] (cf. §6). The grammar for a previous version of SNePS (SNePS-79) that was capable of parsing a reasonable fragment of English directly into SNePS network representations needs to be updated for SNePS-2.1, and needs a capacity to add new words to the lexicon when they are encountered. Developing such a capacity means only that a combination of morphological analysis and grammatical constraint on the parse of the sentence will be used to determine the new word's part of speech. This will suffice to allow the parse of sentences that contain the word, and to permit Cassie to decide whether to use her noun-definition algorithm or her verb-definition algorithm when we ask what the word means. (Such a determination does not actually require a word's presence in the lexicon as separate from the KB: Its position in the network can identify its part of speech. But a simple look-up in the lexicon is faster than searching through the network. As long as we are using a system that requires a separate lexicon for parsing, we may as well take advantage of it.) In addition to adapting the SNePS-79 grammar to SNePS-2.1, we especially intend our new grammar to build appropriate representations of causal and hyponym-hypernym [33] relations between clauses.

8.2 Believing Reported Synonyms. Although Cassie does not add her definitions to her KB, she believes *almost* all the elements of the reported definitions. I.e., the class inclusions and assorted properties of nouns and the results of actions are all present in her KB. However, this is not always true of the reported synonyms. If Cassie believes some word to be a synonym of the target noun, that will be reported, but our system also reports other synonyms without building that information into the KB. This means that we can avoid requiring Cassie to withdraw belief in the synonymy of a pair of words as she learns more about their meanings, but it also means that our system cannot be considered identical with Cassie, since it reports information she does not believe.

We intend to address this problem for the sake both of cognitive validity and of reduced computational expense. In our current implementation, all candidate synonyms (those sharing a superclass with the target noun) must themselves become, in a sense, target nouns subject to portions of the noun-definition algorithm. In a KB of any size, the need to partially define many candidate synonyms each time we wish to define one noun could quickly become prohibitively costly. If Cassie already believes that a word is synonymous with the target noun, it is not necessary to define it.

One approach is to subject all candidates to the partial definition process when we first look for synonyms. Those that survive are asserted in the KB to be *possible* synonyms. The next time, those possible synonyms that survive the process are asserted to be *probable* synonyms. The third time a candidate synonym survives having its partial definition compared with the target noun, we assert that it *is* a synonym. Subsequent requests for a definition of the target noun will then simply report the synonym without it having to be defined. If, later, Cassie learns that her belief in the synonymy of two words was erroneous, then she will need to revise its beliefs.

Another approach is to base the decision about whether a term is a possible, probable, or definite synonym on the amount of definitional information present, rather than the number of times a definition has been reported. If we have structural and functional information on both the target noun and the candidate synonym as well as several matching class inclusions, we might assert the candidate to be a synonym immediately. In a case where we lacked much of the significant information, the candidate might be asserted to be only a possible synonym even if we'd asked for a definition of the target noun many times.

Where we have much relevant information, the second of these approaches could save computational resources. Where we have less information, we might be forced to go through the definition of a candidate many times to see whether its status should be upgraded from possible to probable, or from probable to definite synonymy, rather than being done with it after a fixed number of passes. Either approach, however, would be an improvement on our current system.

8.3 Compiling Definitions. It appears that humans don't store compiled definitions [39]. Rather, we link various aspects of meaning to words, and different portions of a word's meaning "come to mind" depending on the context in which they are used. The only time we are apt to store a compiled definition is when we expect a need to produce that definition, as with a student who must memorize definitions for an exam. Our decision not to build definitions into the KB seems, therefore, to be both cognitively valid and practical (since Cassie need not withdraw "belief" in an early stage of a definition once she learns more).

However, every time we ask her for a definition, she must search out the information afresh even if she has just told us the same thing many times in a row. We currently have no mechanism for compiling a definition, or even of keeping track of how many times a definition has been reported. We are considering the possibility of allowing Cassie to come to believe that certain portions of her knowledge constitute a definition. In §8.2, we discussed using either completeness or stability of definition to determine synonymy. It seems likely that we would want both completeness and stability before allowing the system to store a compiled definition. The stability criterion, however, would require the development of some method for keeping track of how often a particular definition has been presented.

8.4 Verb Definitions. Compared with our noun-definition algorithm, our verb-definition algorithm is sketchy. It could be improved by developing synonym-finding procedures and by the use of some scheme for classifying verbs of different types. One such scheme already available is Conceptual Dependency's primitive acts [67]. Another possibility would be to attempt to develop a classification of verbs analogous to Rosch's basic-level, or subordinate categories [66]. If such a classification is possible, the concepts associated with Schank's primitive acts might be seen as superordinate categories, as would English words such as 'move' or 'go'. 'Walk' might be a basic-level verb, while 'amble', 'pace', 'stroll', 'stride', and 'hobble' might be considered subordinate level. If we can establish such a structure for the classification of (at least some types of) verbs, then, just as we defined 'brachet' as a type of dog, we might define 'striding' as a type of walking.

Determining the class to which a particular verb belongs, as well as determining synonyms, will depend largely on our ability to analyze the results of actions and events. To date, we have paid relatively little attention to the flow of a narrative. In defining nouns, the propositional content of individual sentences seems usually to suffice. In defining the verbs 'dress' and 'smite', it was necessary to notice somewhat the sequence of events. Dressing a sword was reported to enable fighting, because such activities were observed to occur immediately before fighting. The definition of 'smite' had to be revised because the sequence of events in the story had Arthur acting after Lot smote him, which would have been impossible had the smiting resulted in his death. To be generally effective at defining verbs, however, Cassie's capacity for inferring causal or enabling relations between actions and events should be expanded.

Looking for textual markers is an obvious step in deducing causality from narrative. 'So ⟨modifier⟩ that' is a proverbial cause/effect marker, as are 'therefore', 'because', and 'since'; to a lesser extent, 'therewith', 'so that', and 'for then … when' are markers of a cause/effect pair, or at least of enablement/effect. Other such markers exist. Our grammar, once in place, should recognize at least some of these, and build the appropriate relations between the propositions of the surrounding clauses. Temporal terms like 'until' can also be clues. If one activity or state of affairs continues until some event occurs (or other state obtains), the new event or state may cause the termination of the previous activity or state. In the case of a terminated activity, it may end because its goal is accomplished, or because the activity is no longer possible. (The related work of Almeida [1], [2] on time in narratives may be of some use here, as would Cohen's work on textual markers [13], [14].

Sequence of action may be considered in terms of whether it suggests a causal or enabling link. In general, an act by one agent followed by the act of a different agent is not apt to be causal or enabling unless a textual marker is present. When the first verb in a sequence is the action of a living agent and the second is not, the first may be causal. In defining 'dress', we used the heuristic that, when the second verb in a time sequence is a second action by the same agent, the first act may enable the second, or it may enable a larger activity of which the second act is a part. The difficulty with this heuristic is that a list of actions may be effectively simultaneous, if they are done to enable something else. In this passage:

> Then Gryflette *took* his horse in great haste, *dressed* his shield, *took* a spear in his hand, and *rode* a fast gallop till he *came* to a fountain. ([43]: 32; italics added.)

the actions listed are sequential, but each action does not enable the next. Rather, the sequence of actions together enables something. Clearly Gryflette is pursuing a plan for accomplishing a goal. What the goal may be isn't clear from this fragment alone: Taking horse and riding enables coming to the fountain, but dressing shield and taking spear and being at the fountain enable something further (presumably combat, considering the usual associations of spears and shields).

To a large extent, inferring causal or enabling relations in narrative is a matter of having sufficient background knowledge, and does not require changes in our definition algorithms. However, determining which verbs (and how many) form a sequence will require some type of discourse analysis that Cassie currently lacks. A better understanding of a story can only improve the understanding of the words in it.

8.5 Research Priorities. Developing the grammar, resolving the question of when to assert synonymy, and improving the automation of SNeBR/SNePSwD so that beliefs are automatically ranked when a contradiction is noticed and revised rules are automatically asserted in the appropriate context rather than simply being built are system-building tasks we intend to pursue in the immediate future. Deciding if and when to store compiled definitions, adding some capacity for discourse analysis, adding a synonym finder for verbs, and—most importantly—devising some method of classifying them in order to improve the verb-definition algorithm, are research tasks we plan to undertake during the first year and a half of the project, with implementation to be initiated simultaneously (as an essential part of the research) and completed in the second year and a half. Other tasks that remain are discussed next.

8.6 Modifiers and Proper Names. The three major types of open-class words are nouns, verbs, and modifiers. We need to abstract definitions for adjectives or adverbs, and consider methods for defining proper names. For some adjectives, e.g., 'humongous', morphological information might be of use ('humongous' might mean "large", since it seems to be a portmanteau of 'huge' and 'tremendous'). Definitions of color terms might be facilitated using the techniques from related work on color recognition and color naming [41]. Proper names, of interest in applications to reading news items, ought to be handleable by extending our noun-definition algorithm (perhaps to include slots for city/state/country, function, occupation, etc.).

8.7 Malformed Input and Misused Words. When an unknown word is encountered in a text, it may be a genuinely unknown word, or it may be a misspelling or other typographical error. Ideally, Cassie would have a facility for recognizing malformed input, so that it would not be necessary to add 'scjool" to the lexicon and attempt to develop a definition for it just because the text says that Jimmy's mother picked him up after scjool. In practice, this is not something we intend to address. If such a typographical error occurs once, Cassie will simply fail to fully understand the sentence in which it occurs. If it occurs repeatedly, Cassie should eventually have enough contextual information to notice its synonymy with the correct spelling.

Our assumption is that all words in a story are used “correctly”. This is not a completely accurate model of human reasoning. We do sometimes believe that an author or speaker has used a word incorrectly. However, since we are taking a descriptive approach to the definition of words and are interested in allowing Cassie to determine what is meant by a word in a given context, we have not considered any means to allow her to come to the conclusion that a word has been misused. If the usage does not agree with previous understanding, our approach is to develop a secondary definition for it. For the texts we have been using, this approach is quite appropriate. At some time in the future, it would be desirable to give Cassie the ability to recognize some common misusages, such as the frequent confusion between ‘imply’ and ‘infer’, so that she does not come to believe that they mean the same thing. For the immediate future, however, we shall continue to focus our efforts on defining words that are used correctly.

8.8 Using Etymological Roots. When encountering a new word, human readers sometimes look for clues to its meaning in the word itself. Some such clues can be very misleading (e.g., ‘brachet’ rhymes with ‘latchet’ so perhaps it refers to some sort of hardware or fastener), but others may be quite relevant, especially those that are etymological. If one has never before encountered ‘theriodont’ but is familiar with ‘orthodontist’ and ‘periodontal’, one can guess that the new word relates to teeth. In some cases, a familiarity with Greek, Latin, or other roots can allow a reasonable interpretation of a new word even if the context is not very informative. If combined with an informative context, such etymological knowledge can permit very rapid acquisition of a word’s meaning. Although Cassie has no mechanism for considering pieces of words, SNePS’s morphological analyzer gives some of this information, and we will explore adapting it for this use.

8.9 Generating Rules from Cases. By the time a human being is aware of many instances of a category, knowledge of the properties of that category (especially of natural kinds) is usually summarized as default information. Our system searches first for rules containing such information when asked for a definition. If a term (e.g., ‘brachet’) has only recently been encountered, such rules may not yet be present, and reference must be made to those specific individuals of which the system has read. This is not a problem when the number of such individuals is small. But if further reading introduces more individuals, with more properties, there ought to be some means by which rules can be generated from individual cases.

As an example, consider the color of brachets. If we (as human readers) encountered various brachets, all of whom were white, we would probably conclude at some point that brachets, in general, tend to be white. Cassie is able to deduce that the function of brachets, in general, is to hunt, and therefore the stative descriptions of the individual brachet are dropped from the definition. But if she reads of many more white brachets (and none, or few, of any other color), it would be appropriate to build a general rule and return “white” to the definition, this time as a general rather than an individual property.

However, suppose a human reads of black, brown, white, and grey brachets. Such a reader might well conclude (though perhaps not consciously) that no single color is typical of brachets and omit all mention of color if asked for a definition. If Cassie had encountered brachets of many colors before being able to deduce anything about function, our current implementation would simply have listed them all. If she knew of many different actions attributed to members of some class, but had no beliefs about the actions typical of the class, again, they would all be listed. This could, hypothetically, go beyond any reasonable limits. There comes a point at which, if no generalization can be drawn about some aspect of meaning, we should probably cease to report on that aspect, rather than continue to add to a list of slot-fillers based on individuals.

8.10 Automating the Determination of Entrenchment. The assignment of `kn_cat`s to propositions is currently performed by a human. Such assignments are based on intuition about how deeply entrenched a belief should be. Dressing is a daily activity, and the word 'dress' is quite familiar to English speakers, whereas the word 'smite' is not commonly encountered in modern English and our certainty that we understand it may be correspondingly less. We, therefore, marked the rule about what it means to smite as less entrenched than the rule about what it means to dress, and used slightly different algorithms for the revision of these rules. How to formalize this intuitive notion of entrenchment remains a very open question.

Part of the difficulty is that, in attempting to model Cassie on adult humans, we must build in the relevant knowledge such an adult would bring to the text. An adult human's sense of how entrenched his or her understanding of a particular word is depends on experiences with that word: how it was acquired, how often it is encountered it or used. It is this type of judgment based on a summary of experience that we attempt to encapsulate in our `kn_cat`s, but without giving the system the actual experiences that would allow it to form such judgments itself.

This problem is related to the problem of generating rules from individual cases in that both require the ability to summarize and (perhaps more importantly) the knowledge, even if not conscious, of when to summarize. Such ability has not been of great importance in our work to date, since we have been concerned only with acquiring or revising the meaning of a given word from the context of a given story. The knowledge of how entrenched a belief is, like all other knowledge the system has before beginning to read a story, is knowledge we have given it.

Summarizing from many experiences, generating rules from individual cases, and determining entrenchment (and eventually forgetting details) are problems of language acquisition, and of AI in general. Although they will probably remain beyond the scope of our research for the forseeable future (except, possibly, for certain types of rule generation), our system should give us a rich environment for developing and testing such theories.

References

[1] Almeida, M.J. (1987), "Reasoning about the Temporal Structure of Narratives," *TR 87-10* (Comp. Sci., SUNY Buffalo).

[2] Almeida, M.J. (1995), "Time in Narratives," in [19].

[3] Asker, L.; Gamback, B.; & Samuelsson, C. (1992), "EBL^2: An Approach to Automatic Lexical Acquisition" *COLING-92*: 1172–1176.

[4] Berwick, R.C. (1983), "Learning Word Meanings from Examples," *IJCAI-83*: 459–461.

[5] Bruder, G.A. (1988), "The Deictic Center and Sentence Interpretation in Natural Narrative" (Psychonomic Soc. Meeting, Chicago).

[6] Bruder, G.A. (1995), "Psychological Evidence that Linguistic Devices Are Used by Readers to Understand Spatial Deixis in Narrative Text," in [19].

[7] Bruder, G.A. & Wiebe, J.M. (1990), "A Psychological Test of an Algorithm for Recognizing Subjectivity in Narrative Text," *Proc. 12th Annual Conf. Cog. Sci. Soc.*: 947–952.

[8] Bruder, G.A., & Wiebe, J.M. (1995), "Recognizing Subjectivity and Identifying Subjective Characters in Third-Person Fictional Narrative," in [19].

[9] Carbonell, J.G., & Hayes, P.J. (1984), "Coping with Extragrammaticality, " *COLING-84*: 433–437.

[10] Carter, D. (1989), "Lexical Acquisition in the Core Language Engine", *Proc. 4th Conf. European Chapter Assoc. Comp. Ling.*: 137–144.

[11] Choi, J. (1993), "Experience-Based Learning in Deductive Reasoning Systems," *TR 93-20* (Comp. Sci., SUNY Buffalo).

[12] Chun, S.A., & Zubin, D. (1995), "Experiential vs. Agentive Constructions in Korean Narrative," in [19].

[13] Cohen, R. (1987), "Analyzing the Structure of Argumentative Discourse," *Comp. Ling.* 13: 1–10.

[14] Cohen, R. (1990), "Implementing a Model for Understanding Goal-Oriented Discourse," *Int'l. J. Intelligent Systems* 5: 1–31.

[15] Costello, A.M.; Bruder, G.A.; Hosenfeld, C.; & Duchan, J.F. (1995), "A Structural Analysis of A Fictional Narrative," in [19].

[16] Cravo, M.R., & Martins, J.P. (1993), "SNePSwD: A Newcomer to the SNePS Family", *JETAI* 5: 135–148.

[17] Daniels, J.H. (1986) "A Psychological Investigation into the Deictic Center," *Proc. 8th Annual Conf. Cog. Sci. Soc.*: 621–626.

[18] Duchan, J.F. (1995), "Preschool Children's Introduction of Characters into Their Oral Stories: Evidence for Deictic Organization of First Narratives," in [19].

[19] Duchan, J.F.; Bruder, G.A.; & Hewitt, L.E. (eds.) (1995), *Deixis in Narrative: A Cognitive Science Perspective* (Erlbaum), forthcoming.

[20] Ehrlich, K. (1995), "Automatic Vocabulary Expansion through Narrative Context," *TR 95-09* (Comp. Sci., SUNY Buffalo).

[21] Elshout-Mohr, M., & van Daalen-Kapteijns, M.M. (1987), "Cognitive Processes in Learning Word Meanings," in M. G. McKeown & M. E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Erlbaum): 53–71.

[22] Galbraith, M. (1989), "What Everybody Knew Versus What Maisie Knew: The Change in Epistemological Perspective from the Prologue to the Opening of Chapter 1 in 'What Maisie Knew' ", *Style* 23: 197–212.

[23] Galbraith, M. (1995), "Deictic Shift Theory and the Poetics of Involvement in Narrative," in [19].

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

[25] Gleitman, L. (1990), "The Structural Sources of Verb Meanings," *Lang. Acquisition* 1: 1–55.

[26] Gleitman, L. (1994), "A Picture is Worth a Thousand Words—But That's the Problem" (talk presented at the 1st Int'l. Summer Inst. in Cog. Sci. (SUNY Buffalo, July 1994); abstract in *Proc. 16th Annual Conf. Cog. Sci. Soc.*: 965).

[27] Granger, R.H. (1977), "Foul-Up: A Program that Figures Out Meanings of Words from Context," *IJCAI-77*: 67–68.

[28] Granger, R.H. (1983), "The NOMAD System: Expectation-Based Detection and Correction of Errors during Understanding of Syntactically and Semantically Ill-Formed Text," *American J. Comp. Ling.* 9: 188–196.

[29] Haas, N., & Hendrix, G. (1983), "Learning by Being Told: Acquiring Knowledge for Information Management," in R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (eds.), *Machine Learning* (Tioga): 405–428.

[30] Hastings, P.M. (1994), "Automatic Acquisition of Word Meaning from Context," Ph.D. diss. (Comp. Sci. & Eng'g., Univ. of Michigan).

[31] Hastings, P.M., & Lytinen, S.L. (1994a), "The Ups and Downs of Lexical Acquisition," *AAAI-94*: 754–759.

[32] Hastings, P.M., & Lytinen, S.L. (1994b), "Objects, Actions, Nouns, and Verbs," *Proc. 16th Annual Conf. Cog. Sci. Soc.*: 397–402.

[33] Hearst, M.A. (1992), "Automatic Acquisition of Hyponyms from Large Text Corpora," *COLING-92*: 539–545.

[34] Hewitt, L.E. (1995), "Reduced Anaphor in Subjective Contexts," in [19].

[35] Hewitt, L.E., & Duchan, J.F. (forthcoming), "Subjectivity in Children's Fictional Narrative," *Topics in Language Disorders*.

[36] Hill, R. (1994), "Issues of Semantics in a Semantic-Network Representation of Belief," *TR 94-11* (Comp. Sci., SUNY Buffalo).

[37] Hosenfeld, C.; Duchan, J.; & Higginbotham, J. (1995), "Deixis in Persuasive Texts Written by Bilinguals of Differing Degrees of Expertise," in [19].

[38] Jackendoff, R. (1987), "On Beyond Zebra: The Relation of Linguistic and Visual Information," *Cognition* 26: 89–114.

[39] Johnson-Laird, P.N. (1987), "The Mental Representation of the Meanings of Words," in A.I. Goldman (ed.), *Readings in Philosophy and Cognitive Science* (MIT, 1993): 561–583.

[40] Kiersey, D.M. (1982), "Word Learning with Hierarchy Guided Inference," *AAAI-82*: 172–178.

[41] Lammens, J. (1994), "A Comp. Model of Color Perception and Color Naming," *TR 94-26* (Comp. Sci., SUNY Buffalo).

[42] Li, N., & Zubin, D.A. (1995), "Perspective Taking in Mandarin Discourse," in [19].

[43] Malory, Sir T. (1470), *Le Morte Darthur*, ed. R.M. Lumiansky (Collier, 1982).

[44] Mark, D.M., & Gould, M.D. (1995), "Wayfinding Directions as Discourse: Verbal Directions in English and Spanish," in [19].

[45] Mark, D.M.; Svorou, S.; & Zubin, D.A. (1987), "Spatial Terms and Spatial Concepts: Geographic, Cognitive, and Linguistic Perspectives," *Proc. Int'l. Symposium Geographic Info. Systems: The Research Agenda* (Crystal City, VA), Vol. 2: 101–112.

[46] Martins, J., & Cravo M.R. (1991), "How to Change Your Mind," *Noûs* 25: 537–551.

[47] Martins, J., & Shapiro, S.C. (1988), "A Model for Belief Revision," *Artif. Intell.* 35: 25–79.

[48] Nakhimovsky, A., & Rapaport, W.J. (1988), "Discontinuities in Narratives," *COLING-88*: 465-470.

[49] Neal, J.G., & Shapiro, S.C. (1987), "Knowledge Based Parsing," in L. Bolc (ed.), *Natural Language Parsing Systems* (Springer-Verlag): 49–92.

[50] Peters, S.L., & Rapaport, W.J. (1990), "Superordinate and Basic Level Categories in Discourse: Memory and Context," *Proc. 12th Annual Conf. Cog. Sci. Soc.*: 157–165.

[51] Peters, S.L., & Shapiro, S.C. (1987a), "A Representation for Natural Category Systems I," *Proc. 9th Annual Conf. Cog. Sci. Soc.*: 379–390.

[52] Peters, S.L., & Shapiro, S.C. (1987b), "A Representation for Natural Category Systems II," *IJCAI-87*: 140–146.

[53] Peters, S.L.; Shapiro, S.C.; & Rapaport, W.J. (1988), "Flexible Natural Language Processing and Roschian Category Theory," *Proc. 10th Annual Conf. Cog. Sci. Soc.*: 125–131.

[54] Pustejovsky, J. (1987), "On the Acquisition of Lexical Entries: The Perceptual Origin of Thematic Relations," *Proc. 25th Annual Meeting Assoc. Comp. Ling.*: 172–178.

[55] Quillian, M.R. (1968), "Semantic Memory," in M. Minsky (ed.) *Semantic Information Processing* (MIT): 227–270.

[56] Quillian, M.R. (1969), "The Teachable Language Comprehender: A Simulation Program and Theory of Language," *Comm. ACM* 12: 459–476.

[57] Quine, W.V.O. (1951), "Two Dogmas of Empiricism," in *From a Logical Point of View, 2/e, rev.* (Harvard, 1980): 20–46.

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

[59] Rapaport, W.J. (1986), "Logical Foundations for Belief Representation," *Cog. Sci.* 10: 371–422.

[60] Rapaport, W.J. (1988a), "Syntactic Semantics: Foundations of Comp. Natural-Language Understanding," in J.H. Fetzer (ed.), *Aspects of Artificial Intelligence* (Kluwer): 81–131.

[61] Rapaport, W.J. (1988b), "To Think or Not to Think," *Noûs* 22: 585–609.

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

[63] Rapaport, W.J. (1995), "Understanding Understanding: Syntactic Semantics and Computational Cognition," in J.E. Tomberlin (ed.), *Philosophical Perspectives, Vol. 9* (Ridgeview), forthcoming.

[64] Rapaport, W.J., & Shapiro, S.C. (1995), "Cognition and Fiction," in [19].

[65] Rapaport, W.J.; Shapiro, S.C.; & Wiebe, J.M. (1986), "Quasi-Indicators, Knowledge Reports, and Discourse," *TR 86-15* (Comp. Sci., SUNY Buffalo); rev. version, "Quasi-Indicators and Knowledge Reports," *Cog. Sci.*, forthcoming.

[66] Rosch, E. (1978), "Principles of Categorization," in E. Rosch & B.B. Lloyd (eds.), *Cognition and Categorization* (Erlbaum): 27–48.

[67] Schank, R.C., & Rieger, C.J. (1974), "Inference and the Computer Understanding of Natural Language," *Artif. Intell.* 5: 373–412

[68] Segal, E.M. (1995a), "Narrative Comprehension and the Role of Deictic Shift Theory," in [19].

[69] Segal, E.M. (1995b), "Stories, Story Worlds, and Narrative Discourse," in [19].

[70] Segal, E.M.; Duchan, J.F.; & Scott, P.J. (1991), "The Role of Interclausal Connectives in Narrative Structuring: Evidence from Adults' Interpretations of Simple Stories," *Discourse Proc.* 14: 27–54.

[71] Shapiro, S.C. (1979), "The SNePS Semantic Network Processing System," in N. Findler (ed.), *Associative Networks* (Academic): 179–203.

[72] Shapiro, S.C. (1982), "Generalized Augmented Transition Network Grammars for Generation from Semantic Networks," *American J. Comp. Ling.* 8: 12–25.

[73] Shapiro, S.C. (1989), "The CASSIE Projects: An Approach to Natural Language Competence," in J.P. Martins & E.M. Morgado (eds.), *Proc. 4th Portugese Conf. A.I. (EPIA-89)*, Lecture Notes in A.I. 390 (Springer-Verlag): 362–380.

[74] Shapiro, S.C. (1991), "Cables, Paths, and 'Subconscious' Reasoning in Propositional Semantic Networks," in J. Sowa (ed.), *Principles of Semantic Networks* (Morgan Kaufmann): 137–156.

[75] Shapiro, S.C., & Martins, J.P. (1990), "Recent Advances and Developments: The SNePS 2.1 Report," in D. Kumar (ed.), *Current Trends in SNePS*, Lecture Notes in A.I. 437 (Springer-Verlag): 1–13.

[76] Shapiro, S.C., & Rapaport, W.J. (1987), "SNePS Considered as a Fully Intensional Propositional Semantic Network," in N. Cercone & G. McCalla (eds.), *The Knowledge Frontier* (Springer-Verlag): 262–315.

[77] Shapiro, S.C., & Rapaport, W.J. (1991), "Models and Minds: Knowledge Representation for Natural-Language Competence," in R. Cummins & J. Pollock (eds.), *Philosophy and AI* (MIT): 215–259.

[78] Shapiro, S.C., & Rapaport, W.J. (1992), "The SNePS Family," *Computers and Mathematics with Applications* 23: 243–275.

[79] Shapiro, S.C., & Rapaport, W.J. (1995), "An Introduction to a Comp. Reader of Narrative," in [19].

[80] Shapiro, S.C., & the SNePS Implementation Group (1994), "SNePS-2.1 User's Manual," *SNeRG Tech. Note* (Comp. Sci., SUNY Buffalo).

[81] Siskind, J.M. (1992), "Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition," Ph.D. diss. (Elec. Eng'g. & Comp. Sci., MIT).

[82] Siskind, J.M. (1994), "Lexical Acquisition in the Presence of Noise and Homonymy," *AAAI-94*: 760–766.

[83] Siskind, J.M. (forthcoming), "A Computational Study of Lexical Acquisition" (`ftp://ftp.cs.toronto.edu/pub/qobi/cog94.ps.Z`).

[84] Srihari, R.K. (1991), "PICTION: A System that Uses Captions to Label Human Faces in Newspaper Photographs," *AAAI-91*: 80–85.

[85] Srihari, R.K. & Rapaport, W.J. (1989), "Extracting Visual Information from Text: Using Captions to Label Human Faces in Newspaper Photographs," *Proc. 11th Annual Conf. Cog. Sci. Soc.*: 364–371.

[86] Sternberg, R.J. (1987), "Most Vocabulary is Learned from Context," in M.G. McKeown & M.E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Erlbaum): 89–105.

[87] Talmy, L. (1995), "A Heuristic Framework for Narrative Structure," in [19].

[88] *Webster's New Intl. Dictionary of the English Language (2/e, Unabridged)* (Merriam, 1937).

[89] Wiebe, J.M. (1990a), "Recognizing Subjective Sentences: A Computational Investigation of Narrative Text," *TR 90-03* (Comp. Sci., SUNY Buffalo).

[90] Wiebe, J.M. (1990b), "Identifying Subjective Characters in Narrative," *COLING-90*: 401–408.

[91] Wiebe, J.M. (1991), "References in Narrative Text," *Noûs* 25: 457–486.

[92] Wiebe, J.M. (1994), "Tracking Point of View in Narrative," *Comp. Ling.* 20: 233–287.

[93] Wiebe, J.M., & Rapaport, W.J. (1986), "Representing *De Re* and *De Dicto* Belief Reports in Discourse and Narrative," *Proc. IEEE* 74: 1405–1413.

[94] Wiebe, J.M., & Rapaport, W.J. (1988), "A Computational Theory of Perspective and Reference in Narrative" *Proc. 26th Annual Meeting Assoc. Comp. Ling.*: 131–138.

[95] Wilkins, D. (1995), "Expanding the Traditional Category of Deictic Elements: Interjections as Deictics," in [19].

[96] Yuhan, A.H. (1991), "Dynamic Computation of Spatial Reference Frames in Narrative Understanding," *TR 91-03* (Comp. Sci., SUNY Buffalo).

[97] Yuhan, A.H., & Shapiro, S.C. (1995), "Computational Representation of Space," in [19].

[98] Zernik, U., & Dyer, M.G. (1987), "The Self-Extending Phrasal Lexicon," *Comp. Ling.* 13: 308–327.

[99] Zubin, D., & Hewitt, L.E. (1995), "The Deictic Center: A Theory of Deixis in Narrative," in [19].