ROLE: CONTEXTUAL VOCABULARY ACQUISITION: From Algorithm to Curriculum

William J. Rapaport
Department of Computer Science & Engineering and Center for Cognitive Science
Michael W. Kibby
Department of Learning & Instruction and Center for Literacy & Reading Instruction
State University of New York at Buffalo, Buffalo, NY 14260

A. PROJECT SUMMARY

1. Since 7/2001, funded by an NSF ROLE pilot project, William J. Rapaport (Computer Science & Engineering) and Michael W. Kibby (Learning & Instruction) have been investigating "contextual vocabulary acquisition" (CVA): active, deliberate acquisition of word meanings from text by reasoning from contextual clues, prior knowledge, language knowledge, and hypotheses developed from prior encounters with the word, but without external sources of help such as dictionaries or people. The **intellectual merit** of the project lies in increasing our understanding (based on case studies—"think-aloud" protocols) of how good readers use CVA to hypothesize a sense for an unknown word encountered in written context, using these observations to extend the computational theory of CVA, developing further a computer program that implements and tests this theory, and creating and evaluating a curriculum (based on the computational theory) to improve students' abilities to use CVA. The **broader impact** is not merely to improve vocabulary acquisition, but also to increase reading comprehension of STEM texts, thereby leading to increased learning, and, by using a "miniature" (but real) example of the scientific method (viz., CVA), to give students a better understanding of the scientific method. The computational and educational strands of the research are fully integrated and jointly serve these goals. The research falls within **Quadrants 2 and 3**.

2. People know the meanings of more words than they are explicitly taught, so they must have learned most of them as a by-product of reading or listening. Some of this is the result of active processes of hypothesizing meanings for unknown words from context. How do readers do this? Most published strategies are quite vague; one simply suggests to "look" and "guess". This vagueness stems from a lack of relevant research on the reasoning and language processes used for CVA. There is no generally accepted cognitive theory of CVA, nor is there an educational curriculum or set of detailed strategies for teaching it. If we knew more about how context operates, had a better theory of CVA, and knew how to teach it, we could more effectively help students identify context cues and know better how to use them.

Artificial-intelligence and computational-linguistic studies of CVA (including ours) have necessarily gone into much more detail on what underlies the unhelpful advice to "guess", since natural-language-processing (NLP) systems must operate on unconstrained input text independently of humans and can't assume a "fixed complete lexicon". But such studies have largely been designed to improve practical NLP systems. Few, if any, have been applied in an educational setting, and virtually all have been ignored in the reading- and vocabulary-education literature. *AI algorithms for CVA can fill in the details that can turn "guessing" into "computing"; these can then be taught to students.*

This multidisciplinary proposal combines basic and applied research to address twin needs: for NLP systems that operate independently of human assistance and for improving students' abilities to read STEM materials. Its theoretical significance comes from the development of an NLP system that does CVA. Its educational significance lies in whether the knowledge gained from this system can be applied to teaching CVA strategies so that students can use them successfully when they encounter hard words in their regular (STEM) reading. The mutual feedback between the development of the computational theory based on the education team's data and of the algorithm-based educational curriculum is also distinctive, making this a true cognitive-science project.

3. The AI team will: continue developing noun, verb, and adjective algorithms using insights from the think-aloud protocols produced by the education team, develop an NL explanation facility for the system, develop an NL input/output system, investigate the use of "internal" context (morphology) for developing definitions, investigate the possibility of using OpenCYC as one source of general background information, and improve the current definition-revision system.

The education team will design and evaluate a CVA curriculum for middle- and secondary-school students designed to increase: reading comprehension, word consciousness, CVA strategies based on the algorithms, meaning vocabulary, assessments of sense of meaning of hard words, and interest in words. The curriculum will teach students to: focus attention on a hard word in a text (word consciousness) and the textual clues to the word's meaning; connect these clues with their language knowledge, reasoning processes, and prior knowledge to hypothesize a contextual

meaning; and assess their confidence in this meaning. The major instructional strategies of our curriculum are: teacher modeling of CVA via think-aloud procedures, teacher scaffolding the class as it uses CVA, think-alouds for CVA in small peer groups, and individual students' independent applications of CVA with teacher monitoring and assessment. Student materials and a teacher's guide will be developed that will emphasize how this method of CVA is an example of the scientific method "in the small", and the curriculum's effectiveness will be studied.

Section C: Project Description

1 Introduction No doubt you have on occasion read some text containing an unfamiliar word, but you were unable or unwilling to find out from a dictionary or another person what it meant. Nevertheless, you might, consciously or not, have figured out a meaning for it. Suppose you didn't, or suppose your hypothesized meaning was wrong. If you never see the word again, it may not matter. But, if the text were about science, technology, engineering, or mathematics (STEM), not understanding the unfamiliar term might seriously hinder your subsequent understanding of the text. If you do see the word again, you will have an opportunity to revise your hypothesis about its meaning. The more times you see it, the better your definition will become. And if your hypothesis development were active ("deliberate"), rather than passive ("incidental"), your command of the new word would be stronger.

"Contextual vocabulary acquisition" (CVA) is the active, deliberate acquisition of word meanings from text by reasoning from contextual clues, prior knowledge, language knowledge, and hypotheses developed from prior encounters (if any) with the word, but without external sources of help such as dictionaries or people.

Since July 2001, funded by an NSF ROLE pilot project,[1] two integrated teams of researchers from the SUNY Buffalo Departments of Computer Science & Engineering (CSE) and of Learning & Instruction (LAI) (and the Centers for Cognitive Science[2] and for Literacy & Reading Instruction[2], respectively) have been jointly investigating CVA. We have been (a) extending and developing computer programs that do CVA for unknown nouns, verbs, and adjectives, (b) collecting and analyzing verbal protocols of good readers who were asked to "think aloud" as they attempted to hypothesize from context a meaning for an unknown word encountered in that text, (c) unifying a disparate literature on CVA from computational linguistics, psychology, reading science, and first- and second-language (L1, L2) acquisition, in order to help develop these algorithms, and (d) using the knowledge gained from the computational CVA system and the think-aloud protocols to lay the foundation to build and evaluate an educational curriculum for enhancing students' abilities to use CVA strategies in their reading of STEM texts. The knowledge gained from verbal protocols of students using CVA techniques feeds back into further development of our computational theory, which in turn furthers our understanding of how context operates—all of which lay a foundation for building a curriculum to teach CVA [67]. The **broader impact** of our research is not only to improve vocabulary acquisition (which might better be done by direct instruction [63]), but also to increase reading comprehension in general and comprehension of STEM materials in particular, thereby leading to increased learning. Moreover, based on our studies of readers using CVA and our analyses of the CVA research literature, it has become obvious to us that the reasoning and thinking processes people apply to CVA parallel those required in applying the scientific method. The computational and educational strands of our research are fully integrated and jointly serve this broader impact. A more descriptive, albeit overly long, title for our project might be: An Educational Curriculum for STEM Reading Comprehension Based on a Computational Theory of Contextual Vocabulary Acquisition Viewed as an Example of the Scientific Method.

The computational theory and algorithms are being developed in CSE's SNePS Research Group (SNeRG).[2] The analysis of think-aloud protocols and the curriculum development are being done in LAI's Center for Literacy and Reading Instruction. This Center provides supervised teaching practicums for graduate students, who conduct diagnostic assessments and provide reading re-mediation, and is a resource for children and their families to obtain reading diagnostic and re-mediation services. The Center's staff regularly conduct and publish research related to the services they provide to children who find reading difficult. The proposed research enlarges the scope of the Center's mission from strict re-mediation to the provision of vocabulary instruction as part of an in-school developmental program or Center enrichment program.

We seek funding to (1) increase our understanding (based on observations of think-aloud protocols) of how good readers use CVA to hypothesize a sense for an unknown word encountered in written context, (2) use these observations to extend our computational theory of CVA, (3) develop further a computer program ("Cassie"; see §3) that implements and tests this theory, and (4) create and evaluate a curriculum (based on the computational theory) to improve students' abilities to use CVA. Our research falls within **Quadrants 2 and 3** of the ROLE Program (fundamental research on behavioral, cognitive, affective and social aspects of human learning; and research on STEM learning in formal and informal educational settings).

2 Intellectual Merit of CVA The **intellectual merit** of our project stems from the twin needs (a) for NLP systems that can operate independently of human assistance and (b) to improve both the teaching of reading and students'

[1]"ROLE Pilot Project: Contextual Vocabulary Acquisition: Development of a Computational Theory and Educational Curriculum", REC-0106338, $200,000, 7/1/01–12/31/02, with no-cost extension to 12/31/03.

[2]Cognitive Science = [wings.buffalo.edu/cogsci/]; Reading Center = [www.readingcenter.buffalo.edu/]; SNeRG = [www.cse.buffalo.edu/sneps].

reading ability (especially in STEM). Hence, this multidisciplinary proposal combines basic and applied research. Its theoretical significance comes from the development of a theory of how people do CVA and the implementation of this theory in an NLP system that does CVA. Its educational significance lies in whether the knowledge gained in developing this system can be applied to teaching CVA strategies to students so that they are able to use them successfully when they encounter hard words in their regular reading of STEM texts. Our project is also distinctive in its proposed use of mutual feedback between the development of the computational theory and the educational curriculum, making this a true cognitive-science project.

2.1 General significance People know the meanings of more words than they are explicitly taught, so they must have learned most of them as a by-product of reading or listening [55,56]. The *average* number of word families (e.g., 'help', 'helps', 'helped', 'helping', 'helper', 'helpless', 'helpful' are one word family) known by high school graduates is estimated at between 45,000 [55] and 60,000 [54]. *Excellent* students who read a great deal may know 120,000 word families [54]. Learning even 45,000 words by age 18 means learning an average of some 2500 words each year; yet no more than 400 words per year are directly taught by teachers [55]—4800 words in 12 years of school. Therefore, around 90% of the words we know and understand are learned from oral or written context. Learning words from context is not a once-in-a-while thing; it averages almost 8 words learned per day [55]. Some of this "incidental" acquisition is the result of conscious, active processes of hypothesizing the meaning of unknown words from context.

How do readers do this? The psychology, L1, and L2 literatures suggest various strategies [2,9,17,43,84,87,91]. But most are quite vague. Clarke and Nation [17, our italics; cf. 57] gives these directions: step 1: "look at the word itself and its surroundings to decide on the part of speech"; step 2: "look at the immediate grammar context of the word, usually within a clause or sentence"; step 3: "look at the wider context of the word usually beyond the level of the clause and often over several sentences" (looking for causal, temporal, categorical information, etc.); step 4: "*guess ... the word* and check ... that the guess is correct". This is hardly a detailed algorithm that could easily be followed by a student: Step 4 is reminiscent of a famous cartoon showing a complicated mathematical formula, in the middle of which occurs the phrase, "then a miracle occurs"! Part of the problem is that, while many authors suggest what contextual clues to look for (step 3, [83] being the most helpful), few, if any, provide specific advice on what to do with them (i.e., what reasoning or other cognitive processes and what prior knowledge should be applied to them). This vagueness in the educational literature stems from a lack of relevant research on the reasoning and language processes used for CVA.[3] There is no generally accepted cognitive theory of CVA, nor is there an educational curriculum or set of strategies for teaching it. If we knew more about how context operates, had a better theory of CVA, and knew how to teach it, we could more effectively help students identify context clues and know better how to use them, leading to larger vocabularies and better reading comprehension.

"Isn't it easier for the reader to look up the unknown word in a dictionary?" Sometimes; but neither all, most, nor even many new words are learned this way: Dictionaries are incomplete and not always available, their definitions can be hard to understand and not always useful [41,53], and people don't always use them. Nor is learning a significant number of words from dictionaries compatible with the research on meaning-vocabulary acquisition, as just discussed. The dictionary simply is not the major source of learning word meanings in elementary, middle, and high schools [56]. We are investigating ways to facilitate readers' natural CVA by developing a rigorous theory of how context operates and creating a systematic, viable curriculum for teaching the use of CVA strategies, based on artificial-intelligence (AI) research and research analyzing the CVA processes used by good readers.

Sometimes, a reader who "learns" an unknown word just learns how to read it; i.e., the reader—a beginning reader—understands the word in listening and uses it in speaking, but is not able to recognize its printed form. We are not concerned with this kind of "sight-vocabulary" acquisition. By 'vocabulary acquisition', we mean "meaning vocabulary"; it is really the acquisition of a concept or "thing" [10] that is difficult and important. Dale [20] offered four levels of knowledge of *words*: (1) "I never saw it before"; (2) "I've heard of it, but I don't know what it means"; (3) "I recognize it in context; it has something to do with ..."; (4) "I know it". But, following [10], Kibby [43] argues that more important than varying levels of knowing a word are the varying degrees of knowing a *concept*. He defines six such levels:

(1) Knowing both concept and word (or words) that signifies it, e.g., 'finger' on a hand; context is not necessary except to clarify which of several meanings a word might have.

(2) Knowing meaning of idiom or familiar phrase, but not of a specific word in it, e.g., 'arch' in 'arch rival'.

[3] And on how context operates: Little (if any) of the *computational* research on the formal notion of reasoning within a context [31,39,47,52,82] is directly relevant to CVA.

(3) Knowing concept and another word (or words) that signifies it, but not the specific word used in the text to signify it, e.g., readers who know the word 'finger', but not 'digit'; context is most likely to be very helpful.

(4) Knowing concept, but no word for it (context could be helpful); e.g., 'philtrum' signifies the two lines between upper lip and nose.

(5) Not knowing concept or word for it, but having the prior knowledge needed to learn it quickly. E.g., 'pentimento' describes that portion of an old oil painting painted over by a new one that can be seen when the top layer chips or fades; most readers would not know this word, nor are they likely to have ever seen pentimento in a painting (unlike 'philtrum', where most readers have not seen the word but will have seen the physical thing), but even an unsophisticated reader has the prior knowledge necessary to learn this "pentimento" concept. It would take exceptionally strong context to learn a level-5 concept and word, probably an "instructive" context in which the writer intentionally provided contextual information to "define" it. Many encounters with the word in context might be needed for learning a level-5 concept and word.

(6) Not knowing concept or word for it, and also lacking the prior knowledge required to learn the concept quickly. E.g., 'kurtosis' means the relative flatness or peakedness of a frequency distribution as contrasted to a normal distribution; readers will need to know or learn 'frequency distribution' or 'normal distribution' before learning the meaning of 'kurtosis'. Context will not likely provide a sense for 'kurtosis'; only instructive texts will help a reader learn those things they do not currently know nor have the prior knowledge necessary to learn. It might be possible to learn the "kurtosis" concept from numerous encounters with the word in supportive, but not necessarily instructive, texts.

Levels 3–4 require readers to learn a new word for a concept they already know, while levels 5–6 require readers to learn both a new concept and a new word. These six levels of knowing place parameters on the usefulness of context. Levels-3–4 words may be readily learned by CVA, while more contextual encounters or more instructive texts are needed to gain levels-5–6 words from CVA. Each level of CVA learning is important for both school and society. Learning new words for previously known things (levels 3–4) results in: understanding a larger number of the words used in written (and oral) texts, therefore permitting greater comprehension; greater ease and precision in creating oral and written communication; and better understanding of the nuances or shades of meanings of words.

Learning both concepts and their words (levels 5–6) increases the learner's perception and conception of the world and helps "students expand their ability to perceive similarities, differences, and order within the world" [43]. Learning new concepts and their words is not simply "additional knowledge" or learning a definition. Concept learning requires making ever more refined discriminations of ideas, actions, feelings, and objects; it necessitates organizing the newly learned concept with prior knowledge, which generally means reorganizing existing schemata or cognitive organization. Linus Pauling [60] describes this notion: "If I couldn't find a place for some new thing, then I would change my picture of the world until I understood where it fit."

2.2 Computational significance In contrast to the vague strategies discussed above, AI studies of CVA (including our own [22,23,66]) have necessarily gone into much more detail on what underlies Clarke & Nation's unhelpful advice to "guess" the meaning of the unknown word, because natural-language-processing (NLP) systems operating independently of human intervention must not break down when encountering unknown expressions in unconstrained input text, where no assumption of a "fixed complete lexicon" can be made [94]: It could not be manually encoded, nor could it contain new words or new meanings given to old words in new contexts. But such AI studies have largely been designed to improve practical NLP systems. Few, if any, have been applied in an educational setting, and virtually all have ignored, and been ignored in, the reading- and vocabulary-education literature.[4] Moreover, in most of these, the unknown word is assumed to be a synonym for a known word with a known concept. This is especially the case for the large body of research on "word-sense disambiguation", in which the task is to decide which of several possible meanings is the intended one in a given context [38]. As linguist Ellen Prince has suggested (in conversation), that task is similar to a multiple-choice test, whereas CVA is more like an essay test: With our method, it is possible for the system to learn a new concept, at least at Level 5.

Previous computational theories of CVA either (1) use online dictionaries or human informants [94], whereas ours does not; (2) only map unknown terms onto known concepts [35,36], whereas our system is capable of concept *formation*, i.e., *constructing*, rather than merely *finding*, meanings of unknown terms; (3) are concerned with "correct" word meanings [35,36], whereas our goal is reading comprehension, which can succeed with a less-than-fully-accurate definition (see §3); or (4) are concerned with modeling childhood L1-acquisition [81], whereas our focus is on relatively mature readers who already know a large part of their language and are (merely) expanding their vocabulary.

[4] With the possible exception of [11]. See, e.g., [4,5,8,13,16,30,32,35,36,37,44,64,94]. For discussion, see [66]. For a more complete list, see [65], developed as part of the pilot project.

2.3 Educational significance As with a computer, a student who reads an unfamiliar term in a text should be able to figure out what it means, at least to the extent of being able to comprehend the rest of the text, if not to give a formal, precise definition. Helping students with this has been on the "to do" list of teachers for decades. Newly revised educational standards for STEM (as well as language arts and social studies) call for students to have a greater command of concepts and the words that signify them. Since these concepts and words cannot all be taught directly in the classroom, it is important that we not only devote more instructional time in school to teaching CVA, but also gain more knowledge about what context is and how it operates. All textbooks in teaching reading or content areas (including STEM) include admonitions and suggestions to help students in CVA. But reading and content-area experts fall short in specifying: (a) the forms of context, and the helpfulness of a given context, available for CVA; (b) how often a word must be encountered to be confidently learned; (c) how, and how much, prior knowledge can augment available context; (d) the specific reasoning processes that must be applied to these contexts and prior knowledge; and (e) how to evaluate one's confidence in the hypothesized meaning.

The specific context clues that have been recommended for instruction are either too general [2], too specific to experimental conditions [87], too vague [84], or do not result from scientific study of actual CVA, but from logical analyses of language usage and traditional notions about context clues (e.g., most of the L2 and reading literature). There has been little systematic study of good readers using successful CVA processes when encountering unknown words in actual reading (the most often-cited are either old and limited [2] or concentrate on readers who *fail* at CVA [90]). How do readers integrate context, prior knowledge, and reasoning processes in CVA? How much context or how many context clues are needed for a word to be confidently learned or even merely understood sufficiently for general comprehension? Context can be both facilitating [27] as well as misleading [71], but what makes a context helpful, not helpful, or misleading for different kinds of concepts?

There are broader implications than just classroom learning and learning standards, however. Students learn a great deal of STEM from reading websites, trade books (i.e., books that are not textbooks), general-interest children's magazines (e.g., *Highlights*, *Cricket*, *Spider*, *Weekly Reader*), and science-oriented children's magazines (e.g., *Spinner*, *TimeKids*, *Science Scholastic*, *Quantum*). If students are better able to do CVA, then more STEM will be learned when students are independently engaged in such extra-curricular reading. If schools are more effective in teaching CVA, then students will heighten their interest and motivation in, and knowledge of, STEM.

2.4 From algorithm to curriculum *AI algorithms for CVA (ours in particular!) can fill in the details that can turn "guessing" the meaning of unknown words from context into detailed instructions for constructing (i.e., "computing") their meanings; these details can then be taught to students*.

We know of no other *computational-theory-based curricula* like ours. Sternberg et al.'s computer-based *tutoring* system for CVA [84] was never built;[5] in any case, ours is not a computerized tutoring system, but a classroom curriculum developed out of the insights gained from computer algorithms. Aist [1] "focus[es] on developing *computerized methods* to help children learn (as opposed to developing a computer program that learns)", nor is he devising a classroom curriculum (as we are); he is "adding extra information to textual contexts",[5] whereas we are investigating the use of *natural* texts and basing our curricular research on case studies with students actually doing CVA.

Is *teaching humans* how to learn the same as *programming computers* to learn? We view AI as "computational psychology", the study of human cognition using computational techniques; a good computational-psychology computer program will simulate some human cognitive task in a way that is faithful to human performance, with the same failures as well as successes, telling us something about the human mind [74].

Our goal is not to teach people to "think like computers", but to explicate methods for CVA. Clarke & Nation's vague strategy discussed above is *not* a caricature; it is the actual recommendation of respected writers in the field of meaning-vocabulary acquisition! But neither is it an algorithm—i.e., an explicit, step-by-step procedure for solving a problem. Our goal is to "teach" (i.e., program) a computer to do the "educated" guessing—or hypothesizing—that is left vague in their strategy. To do that, we must determine what information is needed and what inference methods must be supplied, and we must spell this all out in enough detail so that "even" a computer could do it [69]. Thus, we believe that the various phrases that have been used in the literature ('learning', 'acquiring', 'guessing', 'deriving', 'inferring', 'deducing', 'educing' (etc.) the meaning of unknown words from context) can all be replaced by one: *computing* the meaning of unknown words from context.

But that is not all: For once we have such a detailed procedure, we can then actually teach it to people, rather than leave them wondering what to do with all the contextual information that they might have found in steps 1–3 of the

[5] Sternberg, personal communication, 1/26/2000. Aist, personal communication, 11/10/2000; our italics.

above vague strategy—we can teach them what information to look for and what to do with it. This is our final goal. And, we suggest, it is precisely what is involved in general reading comprehension.

Thus, it is not our intent to use a computer to teach students, but to apply the knowledge gained from programming Cassie to the creation of instructional methods (for regular classroom milieus) to teach *students* to do CVA. We are not building a teaching machine, but rather teaching a machine, to see if what we learn in teaching it can help us to teach students better.

2.5 CVA and reading comprehension The desirability of improving students' reading-comprehension abilities, especially in STEM, should be obvious. Vocabulary knowledge is highly correlated with reading comprehension [56]. There is empirical evidence for this, but it should not be surprising: The more words whose meanings you know, the better you can understand what you read; conversely, the better you understand a text, the more of its words you know. Thus, improving vocabulary knowledge in general (and CVA in particular) should improve reading comprehension. But this only occurs when the vocabulary acquisition is knowledge-based and integrated into the reader's prior knowledge; it does not occur significantly from rote definition-learning [6,56].

Deliberate CVA can accomplish three major tasks: (1) It can, obviously, increase one's vocabulary. It does this in an *active* way, by having the reader think carefully and deeply about the passage and the unknown word's role in it, and make inferences (deductive, inductive, and abductive[6] about the information in the passage, using the reader's prior knowledge as an extra source of information.

(2) This immediately leads to better comprehension of the passage, for CVA *is* knowledge-based and integrative. Indeed, the end result changes the reader's conceptualization of the world: Both the new word and the new information in the passage are incorporated and integrated into the reader's knowledge base (KB).

During our pilot project, when trying to convert our algorithms into an educational curriculum, we faced an interesting problem. Our computer system has a perfect memory and is a perfect reasoner: Given some prior knowledge and some information in the passage being read, it immediately draws whatever conclusions can be drawn. Real people don't. Even clever graduate students (who can be assumed to be good readers [2]), when reading that a knight picked up a "brachet" and rode away with it, often fail to infer that brachets (whatever they are) must be small enough to be picked up. But our system never fails to infer this. Human readers who miss this inference the first time get it when asked, "How big is a brachet?". So they *can* draw the inference if the relevant information is made explicit. How can we bridge this gap between the computer's perfect memory and human readers' memory limitations?

One answer, we believe, lies in using think-aloud protocols. Our *research methodology* of think-aloud protocols aids in algorithm development and (thereby) curriculum-*content* development. In addition, as a *curriculum method* (a more directive version of research protocols), they can assist the reader in identifying and using relevant prior knowledge, in making appropriate inferences, and in integrating new knowledge from the text with prior knowledge. Reading actively in this way increases comprehension.

(3) One of the goals of education should be to instill in students the knowledge of how to learn on one's own and the confidence and life-long desire to use that knowledge. Often, there are no ultimate authorities or experts to consult when one has a problem to solve or a question to answer [61]. This is just as true in STEM (where there are no "answers in the back of the book") as it is when one comes across an unknown word while reading (and there is no dictionary at hand, or no one else who knows the word). The skills required for CVA are not only useful for helping one read (and hence learn) on one's own, but are also among those most useful in STEM: Finding clues or evidence (among the context surrounding the unknown word) and using reasoning to integrate them with prior knowledge and previous hypotheses to hypothesize a (new) meaning for the unknown word is a wonderful model "in the small" of the scientific method of hypothesis formation, testing, and revision, as well as a useful tool for learning on one's own.

3 Our Computational Theory of CVA What is needed (and what we have been working on) is a general method that shows how CVA can be done and is explicit enough to be taught to human readers. Such a theory is best expressed algorithmically, for then the methods are made fully explicit and can be tested computationally. Although this does not solve the problem of how humans actually do CVA, it does provide testable ideas of how they might do it. And it certainly provides ideas for how they could do it and, hence, how it might be taught.

The computational aspects of our project build upon our previous work on the development of a computational CVA theory and software agent called 'Cassie' [22,23,66]. Cassie consists of the SNePS-2.6 semantic-network knowledge-representation and reasoning system [75–80] and a KB of prior knowledge that a reader (e.g., Cassie) brings to the text containing the unknown term. Currently, the KB is hand-coded: Since it represents Cassie's prior

[6] I.e., inference to the best explanation. (It can also be called "educated guessing"; cf. §2.4)

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.

Each node in a SNePS network represents a concept (possibly built of other concepts), linked by labeled arcs. All information, including propositions, is represented by nodes, and propositions about propositions can be represented without limit. Arcs form the underlying syntactic structure of the network. Paths of arcs can be defined, allowing for path-based inference (generalized property inheritance). There is a 1–1 correspondence between nodes and represented concepts; this uniqueness principle guarantees that nodes will be shared whenever possible and allows nodes to represent concepts, propositions, properties, and such objects of thought as fictional entities, non-existents, and impossible objects [77,78]. This wide representational ability is especially appropriate for CVA from arbitrary texts, whose subject matter could range from factual science to science fiction. After all, objects about which we can think, speak, or write need not exist; this includes not only unicorns, but black holes, bosons, and other theoretical entities of contemporary STEM.

Cassie's input consists in part of textual information, parsed and incorporated directly into her KB. Asking Cassie "What does [word] mean?" initiates a deductive search of the KB, now consisting of a semantic network of prior knowledge plus information from the text, all marked with its "degree" of immunity from revision (roughly, a measure of the trustworthiness of the source of information). SNePS's inference package allows rules for both deductive and default reasoning. In the presence of a contradiction, the SNeBR belief-revision package allows for the removal of one or more of the propositions from which the contradiction was derived, as well as the conclusions that depended on it [50]. This mechanism is used to revise definitions (hypotheses) that are inconsistent with a word's current use (data). 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 [19,51]. SNePS also has an English lexicon, morphological analyzer/synthesizer, and a generalized augmented-transition-network (ATN) parser-generator that, rather than building an intermediate parse tree, can translate the input English directly into the propositional semantic-network KR system [72,73,80].

Output consists of a report of Cassie's current definition of the word in its context (or answers to other queries). We take *the* (conceptual) meaning of a word (as understood by a cognitive agent) to be the position of that word in a single (unified), highly interconnected network of words, propositions, and other concepts, consisting of (in our case) the reader's prior knowledge integrated with his or her mental model of the text being read. In this (idiolectic) sense, the meaning of a word for a cognitive agent is determined by idiosyncratic experience with it. Contextual meaning as thus described includes a word's relation to every concept in the agent's mind, which is too unwieldy to be of much use. But a word's dictionary definition usually contains less information than that. Not all concepts in a network are equally salient to a dictionary-style definition of a word. In the attempt to understand and be understood, people abstract certain conventional information about words to serve as its definition. To limit the connections used to produce the definition, our algorithms for hypothesizing a definition operate by deductively searching the network for information appropriate to a dictionary-like definition (see [22,66] for the original algorithms themselves). Our system produces contextual definitions; i.e., we are *not* proposing a system that develops "correct" definitions (cf. the theory of "preference semantics" [93]). Rather, our system develops dictionary-like definitions that enable the reader to continue with the task of understanding the text.

We have experimented with words that are (a) novel, albeit naming concepts subsequently discovered to be familiar (a "brachet" is a kind of hunting dog), (b) familiar but misunderstood (does 'smite' mean to kill by hitting hard, or merely to hit hard?), and (c) familiar but used in a conceptually new sense (surely "dressing" a sword does not mean to clothe it!). Our theory is that a meaning for such a word, sufficient for understanding the text, (1) *can* be determined solely from context (including the surrounding text, grammatical information, and the reader's prior knowledge, but no access to external sources of information (dictionaries, humans)), (2) can be *revised* upon further encounters, (3) "*converges*" to a stable, dictionary-like definition if enough context has been provided and there have been enough encounters, but (4) is always subject to further revision. Each encounter yields a definition (a hypothesis about meaning) and provides an opportunity to revise it in light of new evidence. The revision is *unsupervised*: There is no (human) "trainer" and no "error-correction" techniques. Finally, no *subject-matter-(or "domain")-specific* prior knowledge is required for the development and revision of the hypothesized definition (with the exception of the word's part of speech). The domain-independence of our system can make it more difficult to develop a good definition quickly, but is intended to model the typical reader of an arbitrary text. Clearly, the more prior knowledge, including specialized knowledge, that the reader brings to the text, the more efficiently the unknown term can be learned.

Cassie was provided with prior knowledge for understanding the King Arthur stories [49]. In one test, when presented with passages involving the unknown noun ‘brachet’, Cassie was able to hypothesize that a brachet was a dog whose function is to hunt and that can bay and bite. ([89] defines it as “a hound that hunts by the scent”.) 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, . . .), 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 prior KB, together with the information she had read in, or inferred from, the text thus far, for information about (1) class inclusions (especially in a basic-level category), (2) general functions of brachets (if known; else those of individuals), (3) the general structure of brachets (or 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 CVA [84,87]. 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 does not), 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.

Our verb-definition algorithm is not yet as elaborate; we currently report its predicate structure, a categorization of its arguments, and any causal or enablement information we can find. We have, however, tested our definition-*revision* procedure on verbs: Cassie was told that ‘to smite’ meant “to kill by hitting hard” (a mistaken belief actually held by one of us before reading [49]). Passages in which various characters were smitten but then continued to act triggered SNeBR, which identified several possible “culprit” propositions in the KB to remove in order to block inconsistencies. Cassie 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 proposed research (see §5.1). Another case is exemplified by ‘to dress’, which Cassie antecedently understood to mean “to clothe”, a well-entrenched meaning that 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 clothe *or* to prepare for battle.

4 Pilot-Project Accomplishments Our pilot project opened new lines of inquiry and demonstrated that the AI researchers and reading scientists could fruitfully work together [67]. Our goals are to (1) make our computational CVA system more robust: to improve and create algorithms for inferring meanings for unknown nouns, verbs, and adjectives from context together with prior knowledge as well as grammatical, morphological, and etymological information, based in part on the strategies gleaned from the think-aloud protocols; (2) construct a theory of how context operates for real readers; (3) develop and test an educational curriculum for teaching CVA, based on both the computational theory as well as the cognitive processes identified in the think-aloud segment of the study; and (4) integrate these tasks by using the computational theory and the educational curriculum to help develop each other.

During the pilot project, the AI team concentrated on revising and adapting the definition algorithms for use by the education team. The education team concentrated on using the computational system to begin the curriculum development, as well as providing feedback to the AI team (based on good readers’ actual CVA techniques) in order to improve and further develop the algorithms. Thus, we are using Cassie to help us learn how to teach humans, and using humans to help us learn how to teach Cassie.

(1) We identified common reading passages, primarily from STEM literature, that we both work on. This, together with our weekly joint research meetings (in addition to more frequent individual meetings) is the focus of the “synergy” between the two teams. We started with the texts that the AI team had already developed algorithms for [22]. To these, we have added texts from STEM textbooks and popular science writing such as is found in *Scientific American* or daily newspapers. The ability to read and understand such texts is an important aspect of STEM literacy, and such writing is more likely to require CVA skills than STEM textbooks (which often give detailed and specific definitions; however, cf. our comments (§2.1) about the limited usefulness of such dictionary definitions).

(2) Because of the lack of knowledge of how context can be used (§2.3), a major strategy of the research conducted by the reading scientists is to study good readers as they attempt to determine the sense of an unknown word encountered while reading. Each is asked to read a “text-set” of passages, each passage containing the same unknown word. After each encounter with the unknown word, they “think aloud” as they apply reasoning strategies, recall prior

knowledge, apply language knowledge, and recollect prior encounters with the word in their attempts to determine what it means. The knowledge gained from these think-aloud protocols in conjunction with knowledge gained from the processes of writing algorithms for Cassie and the information Cassie provides on how she derived word meanings is used in two significant manners: (1) We apply this knowledge to aspects of our developing theory of CVA. (2) We are building a new educational curriculum designed to facilitate students' CVA strategies in the reading of STEM texts.

The education team has collected approximately 100 think-aloud protocols from about a dozen good readers. The texts read focused on STEM vocabulary or words that would be found in STEM texts, e.g., 'estuary', 'proximity', 'proliferate'. Using Web search tools, between 9 and 13 passages were selected for each word. The texts came from a variety of sources, including STEM textbooks, nature and science magazines, general magazines, brochures, government manuals, and websites. In each of the passages for each target word, that word was used between 1 and 3 times; occasionally, an inflected or derivational form of the word was included in one or more of the passages. The passages were selected to represent a variety of reading sources, paralleling the somewhat idiosyncratic nature of encounters with unknown words. The passages within the text-sets were randomly ordered.

Subjects read the text-set one passage at a time. After reading each passage, they were asked to state what they thought the unknown word might mean and to describe their thinking as they came to that meaning. After reading one passage in the set and thinking aloud, they then read the next passage and thought aloud again. On all readings after the first, we ask them how their hypothesis about the word's meaning changed from the previous passage. This continued even after the readers had derived a correct or nearly correct definition, to determine if further encounters with the word reinforced their thinking or changed it.

These protocols have been transcribed, and the analysis of them has just begun. We used protocol-analysis techniques to set the groundwork for collecting and analyzing our data [24,62]. The first stage of this analysis was to code the texts used by identifying the available context clues; we started by using previously published lists of such clues [2,3,21,22,83]. We are presently coding (a) the "correctness" of the hypothesized sense of the word's meaning, (b) the reasoning processes used, (c) prior knowledge used, (d) language knowledge and language clues applied, and (e) words and text strings readers reported using as clues. Currently we have 7 levels of "correctness"; these are still subjects of discussion and revision (cf. §§2.2,3). We began our coding of the reasoning processes using Sternberg's [83] selective encoding, selective combination, and selective comparison reasoning or knowledge-acquisition processes; but we have already added three other processes. We still have much work to complete in exploring the reasoning processes that the subjects applied (whether or not they reported them). We have not yet begun to code the prior knowledge or the change in hypothesized sense of the word over encounters (i.e., changes from the first few passages read to the latter passages read), and will be working on these in the Spring 2003 term.

If our think-aloud methods were for an experiment or case study of CVA, our methods would suffer two major limitations. First, when readers confront unknown words, they are more likely to skip the word than to try to figure out its meaning, especially if the immediate text is understood without knowledge of that word or in situations where excellent comprehension is not a major requirement (reading a novel or magazine article) [33,34]. Thus, stopping readers to require reports on their thoughts about the unknown word creates a not totally normal situation. Secondly, our readers encountered the same word in 9–13 different and consecutively read texts within a 20–40-minute time frame; this is most atypical of normal encounters with unknown words. Often, we may see them only a few times, and with months or years between encounters.

These are not significant limitations in our research, however, as we are trying to maximize each reader's use of context to observe (hear) all the clues, reasoning processes, prior knowledge, language knowledge, and previous knowledge of the word as each reader puts them to use. We want to see what good readers can do when pushed (by being asked to think aloud) to process unknown words frequently (by reading consecutive passages). Our goal here was to learn what strategies are available and could be used, and which strategies good readers use that seem to work. With this knowledge, the education team should be able to (a) provide information to the AI team about what is a text clue for word meanings and the cognitive processes that readers actually use in CVA, (b) gain further information from the AI team that refines, further explains, or extends our think-aloud findings, which, in conjunction with (a), will (c) build a theoretical and methodological foundation for developing a curriculum that could be used to help students better comprehend and learn from STEM texts by improving their CVA abilities. This curriculum will provide theory and background information for teachers, an instructional text for learners, and text-sets for students to learn, practice, and apply CVA processes (see §5.2).

(3) The AI team developed knowledge representations of both the texts and the prior knowledge necessary for understanding and reasoning about them, tested the current system on the unknown words in them, and revised and

updated the original noun- and verb-definition algorithms, making them more efficient, more structured, and more robust; we also tested them on several new words taken from STEM materials. Many of the improvements derived from the student protocols. We also began to develop an adjective-definition algorithm.

(4) Together, the two teams began preparing a literature review (based on [65]) of all the CVA disciplines.

5 Proposed Research We will continue the two-way flow of research results between the tasks of improving the definition algorithms and converting them into a curriculum.

5.1 Algorithm development (1) We will continue making our algorithms more "robust", in part by incorporating insights from the think-aloud protocols. We plan to make the noun algorithm more sensitive to a wider variety of information found in the text and Cassie's prior knowledge. For instance, many passages implicitly define new terms without labeling these as "definitions"—our algorithms must use this information, but should not completely rely on it: Other information may help produce a clearer definition or one more integrated with Cassie's prior knowledge (and thus more likely to be retained, in the case of humans whose memories are less perfect than Cassie's). We also plan to improve Cassie's ability to do inductive reasoning (e.g., to be able to generalize from 'dress' meaning either "clothe" or else "prepare for battle" to meaning "prepare" in general).

(2) The current verb algorithm needs to be extended to take into account the wide and often unpredictable variety of "subcategorization" (or "case") structures. We also plan to make use of verb categories (e.g., 'walk', 'run', 'saunter', etc., are all kinds of movements-by-foot) [48,70].

(3) We will continue developing an adjective algorithm. Adjectives are much less susceptible to CVA than nouns or verbs. (Does the neologism 'foo' in the phrase 'the foo car' mean a shape, a color, a size? Is it temporal, like 'new' or 'old'? Could it be like 'toy', which would imply that a foo car isn't even a car?) But some adjectives can occur in contexts that enable useful CVA (e.g., if Fred is "taciturn", unlike his brother who is talkative, we should be able to infer at least that 'taciturn' might mean "not talkative").

(4) All of our algorithms now include "tracing" features (developed during the pilot project). The primary purpose of these is for debugging, but they should also be adaptable for finding out how Cassie inferred her definitions. Thus, we should be able to develop an *explanation facility* for Cassie. This would not necessarily be part of our proposed curriculum, but merely an optional application of the technology that could be used if students have trouble hypothesizing a meaning for an unknown word (or merely want to check to see if their hypothesis is acceptable); they could ask Cassie for an explanation of how she developed her hypothesis (or, better, for guidance on how the *student* could develop one).

(5) Another major task is the development of a method for NL (English) input. Our current methodology has been to concentrate on the KR component, namely, the semantic-network representation of the text containing the unknown word, integrated with the prior knowledge needed, since that is what the algorithms deal with. However, we need to automate the process of representing the text in the KR system. This is best done using a semantic parser that can transduce NL text into a KR. Currently, SNePS does this using an ATN parser, but work is underway to use the LKB computational grammar-development system [18], which has wider coverage than our "home-grown" ATN grammars.

(6) One other advantage of using a grammar with a morphological analyzer (instead of hand-coding the text) is that we can also begin to investigate the use of "internal" context [84], i.e., the morphological and etymological structure of the unknown word, for developing definitions. One reason we began our research using "external" context is that using morphological information by itself is not sufficient. External context of the sort we are investigating is needed as a constraint (e.g., morphology alone might suggest that 'inflammable' means "not flammable", but external context might indicate the opposite).

(7) A generation grammar that would convert the KR to English (also part of SNePS [72]) would enable us to express definitions more completely and more naturally than our current style of definitions, which is simply a slot-and-filler frame, often containing cryptic information.

(8) We also plan to investigate the possibility of using the OpenCYC KB to serve as one source of general background information. OpenCYC is an open-source version of the CYC knowledge "server", containing a large collection of "commonsense" information and reasoning rules.[7] If usable and compatible with SNePS, it would greatly facilitate the development and encoding of prior knowledge.

(9) Finally, we plan to replace SNePSwD by AutoBR [76]. SNePSwD was used to partially automate the revision of hypothesized definitions. AutoBR is a newer version of SNePS's belief-revision system that should allow more automated updating of hypotheses.

[7] OpenCYC = [www.opencyc.org/]; CYC = [www.cyc.com/].

5.2 Curriculum development and evaluation Our pilot project helped us see progress and accomplishments, as well as mistakes and omissions. These insights helped us clarify the specific tasks we need to accomplish. We made significant progress (a) developing procedures for collecting think-aloud protocols as students apply CVA, (b) identifying specific context clues readers use for CVA, (c) scaling the reader's sense of an unknown word's meaning, and (d) broadening (significantly beyond extant published findings) observations on CVA reasoning processes. We still have much to do, however, before presenting a CVA curriculum. (1) We will complete our analysis of the think-aloud protocols we have collected from good readers. Our analyses have so far revealed numerous important reasoning and cognitive processes used during CVA by good readers (including frequently occurring non-productive strategies). In conjunction with the definition algorithms, these think-aloud findings will significantly influence not only the theory of CVA, but also instructional methods for CVA. (2) We will build and evaluate a CVA curriculum designed to help students increase knowledge of word meanings, leading to greater STEM reading comprehension and content learning. The curriculum will be based on our algorithms and will use teacher-modeled protocols that will be practiced by students with the teacher, in small groups, and alone.

5.2.1 Year 1 goals: (1) Develop further knowledge of the specific reasoning processes readers use for CVA, by (a) further think-aloud-protocol collection, (b) using Cassie's algorithms, and (c) broadening our literature review to include comprehension monitoring, strategic reading comprehension, cognitive science, and critical thinking. (2) Determine the effects on CVA processes of variation in readers' depth of knowledge of the concept the hard word signifies (3) Determine the effects on CVA processes of variation in the grammatical category of hard words. (4) Continue planning and designing the curriculum.

Year 1 tasks to accomplish these goals: (1) We will continue analyzing think-aloud protocols to study the interaction of context, reasoning, language, and prior knowledge in CVA. We will continue collecting think-aloud protocols from readers as they apply CVA strategies, but hone our data collection five ways: (a) Collect protocols from both skilled and less-skilled readers to sharpen our focus on productive CVA reasoning processes. (b) Help readers identify important text segments to determine the reasoning that students are/are not able to deploy when directed to do so. This may be a productive teaching technique in our curriculum. (c) Explore the value of retrospective think-alouds (students listen to tapes of their think-alouds while re-reading text). (d) Collect protocols on verbs and adjectives. (e) Collecting and transcribing protocols is time-consuming and limits the number of subjects from whom data may be collected. We will experiment with the techniques and value of readers logging their CVA reasoning, prior knowledge, and language processes in a "lab notebook". (2) In the spring of Year 1, we will actively begin to build our CVA curriculum.

5.2.2 Year 2 goals: We will conduct quantitative experimental research studies to evaluate the draft version of our curriculum. On the basis of the results of these experiments, we will revise the draft version into the "beta" version, which, in Year 3, will be experimentally tested head-to-head with other accepted methods of teaching context clues and vocabulary.

Year 2 tasks: (1) Completing draft instructional materials, (2) designing and validating instructional and experimental tests, (3) conducting an experimental evaluation study of the effects of our curriculum, and (4) conducting experiments to study variables associated with success in using CVA.

Curriculum rationale and goals: In a *Blondie* comic strip years ago, Dagwood's young neighbor Elmo asked why school taught so many different words for things he already knew and for which he already had a perfectly good word: "I know 'fast'; why do I have to learn 'quick', 'rapid', and 'speedy'?". Little did Elmo realize that when a reader does not know the meaning of a word in a text, comprehension may fail. So, if Elmo knows only the word 'fast' when reading a text using 'quick', 'rapid', or 'speedy', he may not comprehend—unless the context leads him to associate these words with the concept "fast" he already knows. Not every encounter with an unknown word leads to poor comprehension, but many do. In cases where the reading is for pleasure, the reader may choose to ignore the word and just read on. But where comprehension is important—STEM reading—readers would be wise not to skip unknown words that affect comprehension. This makes CVA important. Further, learning new things and the words that signify them increases or restructures knowledge; also larger vocabularies are valued socially and for communication.

Our CVA curriculum has 6 goals: to increase (1) reading comprehension, (2) word consciousness, (3) CVA strategies, (4) meaning vocabulary, (5) assessments of hypothesized meanings of hard words, and (6) interest in words.

Concepts (things): Several factors affect CVA: the reader's (1) cognitive awareness of encountering an unknown word (word consciousness), (2) ability to find supportive context clues in the passage, and (3) ability to connect reasoning, language facility, and prior knowledge to these clues. Prior knowledge includes how much the reader knows about the thing signaled by the unknown word (§2.1). Context is best for helping readers learn words they do not know for things they do know. Thus, most words used in a CVA curriculum should signify things students already know. Earlier (§2.1), we defined these as levels 2–4 and some level 5. Like Elmo, one might ask why learning words for things already known is important. An explanatory example: Knowing the two lines between nose and upper lip (the thing) makes it easier to use CVA to learn the meaning of 'philtrum' (the word), and when context allows readers to connect an unknown word to a known thing, text comprehension is likely to increase. Readers may also add a new word to their mental lexicon.

Words: Low-frequency, real words [15] will be used. They will be pre-tested to assure appropriateness for students in the study. Our think-aloud research has found that use of nonsense words does not hinder students' application of context clues or reasoning, so we may occasionally use nonsense words—but only for instruction, not for research evaluation purposes. The words used will signify things that are at levels 2–4. We will also use level-5 concepts, but the supporting context must provide a link between the unknown concept and the prior knowledge required to learn the concept and its word. Level-6 concepts will be used rarely.

Reading passages: A frequent criticism of CVA research is that texts used are contrived, thus lacking in generalizabilty. As in our pilot project, all texts used in teaching and research will be authentic, expository texts from STEM textbooks, magazines, brochures, books, reports, websites, etc. To learn a word, one usually needs to encounter it in more than one text. As in our pilot project, we will use text-sets of 3–8 short passages, each passage in a text-set including the same unknown word.

Instructional components: We will show students how to (1) *focus attention* on a hard word in a text (word consciousness) and the relevant textual clues, (2) *connect* these clues with their language knowledge, reasoning processes, and prior knowledge in order to (3) *hypothesize* a meaning for the unknown word. Students will also learn to (4) *assess* their confidence in this meaning to determine if they need to search elsewhere for a definition. Space prohibits describing each component, so we focus on the major difference between our program and other CVA programs—component (2): connecting reasoning processes, language knowledge, and prior knowledge to context clues.

More than 85 years ago, [86] described reading as a "reasoning process": "understanding a paragraph is like solving a problem ... selecting the right elements and putting them together in the right relations". Most CVA research (e.g., [25,46,85]) has focused on "selecting the right elements" and on practice. In our curriculum, we will identify "the right elements" (e.g., [3,21,22,83]) and provide practice, but more important is helping students "put them together", i.e., connecting these "elements" (context clues) with reasoning strategies, prior knowledge, and language knowledge. It is this focus on *connection* that differentiates our research and teaching methods from almost all previous CVA programs and studies. All four instructional stages of our curriculum will connect reasoning and language to context and prior knowledge.

Reasoning: In our initial analyses of think-aloud protocols (which needs further work), we found several forms of reasoning applied by good readers: visualization; summation; association; clarification; comparison/contrast; questions; predictions; anaphora; and deductive, inductive, and abductive thinking.

Language knowledge: Our pilot project found that readers use certain language knowledge for CVA, e.g., words in a series, question-answer patterns, cause/effect parallelism, and comparison/contrast patterns (cf. [2,3]).

Prior knowledge: To accommodate students' prior knowledge, words in our curriculum will be unknown by students, but they will most likely know the thing the word signals (e.g., two lines above nose/'philtrum').

Major instructional stages: Instruction will progress through four stages in the curriculum. (1) From teacher modeling by think-alouds to (2) teacher scaffolding the class to (3) peer-group think-alouds and discussions to (4) students independently applying CVA strategies with teacher monitoring and assessment.[8]

Teacher modeling by thinking aloud: Unlike teaching the concrete algorithms for long division or the stages of a lunar eclipse, directly teaching someone how to reason is not possible. The closest approximation to teaching someone how to reason is for teachers to *model* such reasoning by *thinking aloud* as they apply CVA strategies. Teacher modeling is effective for enhancing comprehension and comprehension monitoring of older readers [58,92]. In our research review, we found few studies that explicitly used think-aloud techniques for teaching CVA. In [29], the teacher "modeled the process of word-meaning acquisition and guided students through the process" for 5 remedial readers in a 1-1 situation, but no specific description of this modeling or guidance is ever presented. In our curriculum,

[8] As it happens, this follows [61]'s developmental stages.

teachers will model the identification of context clues in a text, but, more importantly, they will model how they connect these clues to reasoning processes, prior knowledge, and language facility to hypothesize possible meanings for unknown words. Further, students will use think-aloud methods themselves in large and small groups.

Scaffolding: 'Scaffolding', a Vygotskian term, means that teachers will assist the whole class in learning CVA strategies. There are two stages to the scaffolding. First, students will independently keep a "lab notebook", logging the CVA strategies they used to develop their hypotheses about the meanings of hard words. Second, using the logs, they will share their reasoning with peers as the teacher helps them develop, evaluate, and test their hypotheses by directing, monitoring, suggesting, and letting discussion flow. Teacher scaffolding will gradually decrease.

Peer groups: Again, students will first independently log CVA strategies in their lab notebooks; then, working in groups of 4–6, they will mutually think aloud as they develop, evaluate, and test their hypotheses. Similar to [26]'s peer-group cues for reading comprehension, specific visual prompts will be provided to help the group focus its attention on context clues and reasoning processes. Peer-group sessions will continue for 3 weeks. Discussions and think-alouds in peer groups during reading comprehension lead not only to better comprehension of the text read and discussed, but students also further develop comprehension strategies [26]. Think-alouds in peer groups should also facilitate CVA. Throughout this stage, peer-groups will assess their confidence in their hypotheses of the word's meaning; this will be done to help students see that they may need to revise their hypotheses. Our curriculum will break new ground in teaching CVA; think-aloud techniques will be used by both teachers and students.

Independent application of context clues and self-assessment: Developing word consciousness requires not just instruction, but repeated practice. In our curriculum, students will apply CVA strategies independently for several weeks, with intermittent peer-group sessions.

Overview of the experimental study: Before describing the materials, tests, dependent variables, and data-analysis techniques of the study, we briefly overview our experimental design and procedures.

Summary of design: Two questions guide this study: (1) Does our curriculum improve CVA and comprehension? (2) Is student reading ability associated with learning in our curriculum? In general, the study uses a 3 x 4 ANOVA design. Reading ability is the first factor: skilled, average, and less-skilled readers (see "Students", below, especially comments regarding low-ability students). Treatments is the second factor: our curriculum, practice only, pre-test/post-test control, and post-test-only control (see "Treatments"). We will construct measures of 8 dependent variables to test the effects of reading ability and treatments (see "Test" and "Dependent Variables" sections). Most of our analyses will be separate 3 x 4 univariate ANOVAs for each dependent variable, but in a secondary analysis to determine the effects of curriculum learning over time, we will group 3 dependent variables for a 2 x 4 x 5 (time) MANOVA (see "Data Analyses"). We will also compare pre- and post-instruction CVA think-aloud protocols from 3 students at each of the reading-skill levels in each treatment group except post-test-only.

Teachers: In Year 2, CVA researchers will teach the curriculum, to gain first-hand experience that will be useful in revising the draft curriculum. In Year 3, regular teachers will teach the curriculum, to establish ecological validity.

Students: Intact, grades 6–8 classrooms will be used in Year 2. The study's focus will be skilled readers with scores on a standardized reading test at $\geq$ 67th percentile and average readers with scores between 37th and 66th percentile. Because we will use intact classrooms, less-skilled readers (who read below the 37th percentile) will be exposed to the curriculum. They will have major word-recognition difficulties that diminish comprehension [28,42]; this in turn interferes with CVA. To assure sufficient comprehension for CVA, Goerss et al. [29] often read texts to their remedial students. After collecting and analyzing the data from our study, it may be found that data from less-skilled readers are not valid and should be excluded. If so, the main design will be a 2 x 4 ANOVA.

Treatments: This experiment will have 4 treatments. (1) <u>Curriculum:</u> Our 10-week curriculum will be based on (a) teachers thinking aloud to model their own CVA strategies, (b) teachers scaffolding the entire class to think aloud while using CVA strategies, (c) peer-group sessions where students in groups of 6 use think-aloud while using CVA strategies, (d) individual students' independent applications of CVA strategies. Curriculum students will also complete an instructional transfer passage and tests each day. (2) <u>Practice only:</u> Each day for 10 weeks, the practice-only group will read the same set of practice and transfer passages used by students in the curriculum, but without teacher modeling, instruction, practice thinking aloud, or small-group work. Independently, they will (a) read each passage in the instructional text-set, (b) use context to hypothesize the hard word's meaning, and (c) complete each day's instructional transfer passage and tests. (3) <u>Pre-test/post-test-only</u> control group will receive no instruction, but complete all pre- and post-tests. (4) <u>Post-test-only</u> control group will complete only the last post-test.

Goal 1. Develop instructional materials: Major materials are a teacher's guide, student text, and reading passages.

Teacher's guide: We will prepare a comprehensive guide with two main sections: (1) important and useful background information on the development and teaching of meaning vocabulary; (2) an instructor's guide to help teachers deploy the curriculum.

Student instructional text: We will prepare a student text that provides an introduction to indirect and direct methods of learning meaning vocabulary, instructional guidelines emphasizing the scientific method of hypothesis development and testing, and STEM passages with unknown words (see next section).

Instructional and transfer passages: Recall, the Year 2 evaluation study is designed (1) to determine if our curriculum has any instructional merit at all and (2) to allow us to test our instructional methods, materials, tasks, and assessments. In Year 3, we will evaluate our curriculum's instructional validity in controlled experiments comparing it head-to-head with other accepted vocabulary-instruction programs.

Instructional passages: There will be 50 text-sets for instructional purposes, one per day. Text-sets will consist of 3–8 passages of 20–200 words, each passage within a set containing at least one instance of the same hard word; hard words will vary in degree of concept difficulty. Passages in text-sets will vary in their sources and topics.

Instructional transfer passages: Concluding each instructional day, students will read a different, short (< 100-word), instructional transfer passage with one hard word varying in concept difficulty from levels 2 to 4, which will not be identified before reading or graphically highlighted in texts. Each passage will be accompanied by 6 tests to measure comprehension and word meaning (see "Goal 2" and "Dependent Variables").

Research transfer passages: To evaluate the effects of the curriculum, students will read a different research transfer passage as a pre-test and as post-tests 1 day, 1 month, and 2 months after instruction. Each research transfer passage will be 700–1,000 words and contain 20 hard words; 18 of the hard words will vary in concept difficulty from levels 2 to 4, 2 words will be level 5. Each passage is accompanied by 6 tests of comprehension and knowledge of the hard words (see "Goal 2").

Goal 2. Develop tests for instructional assessment and program evaluation: For teachers to assess student progress and for us to evaluate the curriculum, 8 tests will be developed. All 8 will be dependent variables in our Years 2 and 3 evaluation studies. The first 6 tests will be integrated into daily instruction: (1) reading comprehension, (2) word consciousness, (3) incidental sense of the meaning of hard word(s), (4) confidence in hypothesized word meaning, (5) deliberate sense of the meaning of hard word(s), and (6) ability to keep a log of reasoning processes, prior knowledge, and language knowledge used to hypothesize a meaning for a hard word. Vocabulary in isolation (7) is a measure of students' knowledge of the meanings of the 180 hard words presented in this study and will be administered 5 weeks before the pre-test and as part of the last post-test. The final dependent variable is (8) the inter-subject convergence on a meaning for the hard word (the degree to which students' theories of the meaning of a hard word converges). Each of these tests is described next:

Test 1. Reading comprehension: The most important reason for CVA is to improve reading comprehension. After reading each day's transfer passage, students will be given 4 multiple-choice or short-answer questions, the answer to one always depending on understanding a phrase, clause, or sentence containing the hard word. For pre/post-test transfer passages, there will be more comprehension items, and the answer to 10 items will always depend on the information in a phrase, clause, or sentence containing the hard word.

Test 2. Word consciousness: A CVA curriculum surely must help students recognize when they encounter words in a text whose meaning they do not know. This is not something that all readers readily do [29,33,34,45,68]. To determine the development during instruction and long-term effects of the curriculum on word consciousness, word consciousness will be tested on the pre-test, the 50 daily transfer passages, and on the 3 follow-up evaluations. Word consciousness will be measured by providing students a list of words that always contains some easy words students have definitely read recently, the hard word(s) from the transfer passage, and—as validity checks—some rare or nonsense words. Students will check all those words they have recently seen in print.

Test 3. Incidental sense of hard word's meaning: Though not the curriculum's purpose, it is certainly expected that students learn some hard words used in instruction. Previous studies estimate that 5%–22% of unknown words encountered during reading are learned incidentally [55]. After reading the transfer passage and without referring to it, students' theories of the meaning of a transfer passage's hard word(s) will be tested one of three ways: (a) multiple-choice tests in which the distracters (the multiple-choice items that are not the desired response) vary in degree of similarity to the target word, (b) multiple-choice item(s) where the student identifies which of 5 complete sentences uses the word accurately, and (c) written definition(s). The multiple-choice vocabulary tests conform to [56]'s 3 levels of distracters: *Lowest-level* distracters denote concepts different and unrelated to target word; part of speech is different. *Intermediate-level* distracters share the same part of speech, but differ significantly from the target word in the concepts they represent. *Highest-level* distracters represent similar or closely related concepts. These tasks vary in difficulty from easy to difficult. On pre/post-tests, both high-level multiple-choice items and written statements of meaning hypothesis will be used. During the duration of the curriculum, there will be a gradual progression from easy (e.g., low-level vocabulary test) to moderate (e.g., high-level or sentence multiple-choice) to most difficult (i.e., writing a theory of the word's meaning).

Test 4. Assessing confidence in theory of meaning of hard word: With CVA instruction, students should become better able to judge how confident they are in their hypothesis of the meaning of a hard word they have encountered. We will measure this confidence after the reading of all transfer passages, on a simple rating scale of 0–5 (0 = no confidence).

Test 5. Deliberate sense of meaning of hard word: Test-3 tests students' incidental learning of a hard word from CVA when the transfer text did not highlight the hard word and students were not told to attend to the word. Test-5 will again test the student's ability to hypothesize the same hard word's meaning, but this time a second copy of the transfer text will highlight the hard word, and students will be directed to write their hypothesis of the word's meaning.

Test 6. Lab-notebook recording of CVA strategies used: After Test 5, students will keep a lab notebook logging the reasoning processes, prior knowledge, and language knowledge used to form their hypothesis of the meaning of the hard word. They will do this for each word on the daily transfer passages, and for a selected set of 5 words on the pre/post-test transfer passages. The dependent variable will be measured by a rating of 0–4 of the depth and quality of the processes stated.

Test 7. Vocabulary in isolation: This test will consist of the 50 hard words from instruction, the 50 from the instructional transfer passages, and the 80 from the pre/post-tests. The high-level multiple-choice format will be used. Number correct will be the measure of the dependent variable. This will be administered once as a pre-test and once as the last post-test.

Test 8. Inter-subject convergence on hypothesized meaning of hard word: We found in our pilot project that when groups of readers read text-sets containing a hard word, many members of the group converged to a similar hypothesis about its meaning. We have not yet developed a way of scoring this dependent variable, but we have observed this convergence in our protocols and in our own attempts to derive hard words at research-team meetings.

Goal 3. Curriculum evaluation:

Dependent variables and procedures: These 8 tests are the dependent variables in our research design. The table provides an outline of the time frame of the experiment (Weeks 1–25). It shows the instructional activities and transfer tests (daily and post-test) for the 4 groups as well as when the tests will be administered to each of the 4 groups over the course of the experiment as measures of the dependent variables.

Week	Curriculum Treatment	Practice Only	Pre/Post	Post Only
1	Test 7: Vocabulary in Isolation	Same	Same	Nothing
6	Pre-test Transfer Passage: Tests 1–6	Same	Same	Nothing
7-16 (Each of 50 Days)	Instruction: Modeling → scaffolding → peer groups → independent; Daily transfer text: Tests 1–4 & (with highlighted text) Tests 5–6	Instruction: None, but complete instructional texts; Daily transfer text: Tests 1–4 & (with highlighted text) Tests 5–6	No instruction, no practice	No instruction, no practice
17	Post-test 1 transfer text: Tests 1–6	Same	Same	Nothing
21	Post-test 2 transfer text: Tests 1–6	Same	Same	Nothing
25	Post-test 3 transfer text: Tests 1–6 Test 7: Vocabulary in Isolation	Same	Same	Same

Data analyses: As described earlier, the main portion of our data analysis will be a 3 (reading ability) x 4 (treatments) ANOVA design. We will use this design to analyze average scores on each of the 8 dependent variables. (If the data from low-ability students are not valid due to inability to read the words of the passage, their data will be omitted, and the main analysis will be a 2 x 4 ANOVA.) The analyses of the pre-test scores will be a 3 x 3 design (not including the post-test-only group, since they will have no pre-test scores). A secondary analysis will be a 3 x 2 x 5 MANOVA with repeated measures on the third factor. The factors included will be reading level, treatments (excluding the 2 control groups in this analysis), and time. The time factor will be the five 2-week segments of the instruction (i.e., weeks 1–2, 3–4, . . . , 9–10) and is a repeated measure. The dependent variables will be reading-comprehension, word-consciousness, and sense-of-word-meaning scores averaged for each of these five 2-week segments. A significant interaction is hypothesized, since it is expected that the rate of learning will be more rapid for the curriculum group than the practice-only group.

Goal 4. Studies of variables associated with curriculum learning: Several variables have been associated with vocabulary learning in past research. Student reading level is one such variable, and we will evaluate reading-level effects in our Year-2 evaluation study. Other variables thought important that we will research in experiments include student vocabulary size, student degree of knowledge of the concept identified by the unknown word, and part of speech of the unknown word. Space precludes a full description of these studies.

5.2.3 Year 3 overview: Year 3, the project's final year, is when we evaluate the validity of our curriculum head-to-head with other accepted vocabulary and CVA methods and programs. These will be controlled, quantitative experiments to see which leads to stronger CVA strategies, greater comprehension, and a better meaning vocabulary. They will be similar in design and conduct to the Year-2 study, but with different and additional instructional treatments.

Goals for Year 3: (1) To revise the draft curriculum in light of the results of Year-2 studies, and write the beta version. (2) To conduct experiments comparing the effectiveness of the curriculum to other context and vocabulary instructional programs. We will conduct experiments comparing the curriculum to both (a) instructional programs presented in previous research studies (e.g., [7,12,14,29,40,59,88]) and (b) commercially-produced vocabulary development programs such as *Word Smart Jr.* and *Wordly Wise*. More than one study will be conducted, comparing the curriculum to other established context and vocabulary-development programs in a variety of school settings. It is anticipated that most studies will be 2-way ANOVAs, with the first factor being reading level and the second being treatment. It is possible that the results of our Year-2 studies of student vocabulary size and student degree of knowing a concept may require us to control these variables in our studies, thus necessitating a more complex 3- or 4-way ANOVA.

6 Summary

We propose to conduct systematic cognitive, computational, and educational studies of CVA, both to understand how context does and does not operate as well as to teach students to be more efficient in their use of it. We have an unusual opportunity here to blend the talents of cognitive computer scientists and reading scientists to explore these issues. Determining how humans do (or *should* do) CVA is a research problem well suited for AI, cognitive science, and reading education. This application of AI will result in new knowledge about how humans do CVA, and make it possible to create better curricula and instructional materials for teaching CVA strategies in reading STEM—in turn increasing reading comprehension and learning. An essential goal of all education is to develop independent, proficient, and creative readers and learners. Most of the knowledge we need to communicate with the world is conveyed by words; therefore, helping students develop lifelong abilities to learn words and concepts from reading is important.

Section D: References Cited

1. Aist, Gregory (2000), "Helping Children Learn Vocabulary during Computer Assisted Oral Reading", *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000; Austin, TX)* (Menlo Park, CA, & Cambridge, MA: AAAI Press/MIT Press): 1100–1101.

2. Ames, Wilber S. (1966), "The Development of a Classification Scheme of Contextual Aids", *Reading Research Quarterly* 2: 57–82.

3. Ames, Wilber S. (1970), "The Use of Classification Schemes in Teaching the Use of Contextual Aids", *Journal of Reading* 14(1): 5–8, 50.

4. Asker, Lars; Gamback, Bjorn; & Samuelsson, Christer (1992), "EBL^2: An Approach to Automatic Lexical Acquisition", *Proceedings of the 15th International Conference on Computational Linguistics (COLING-92, Nantes, France)* (Morristown, NJ: Association for Computational Linguistics): 1172–1176.

5. Barg, Petra, & Walther, Markus (1998), "Processing Unknown Words in HPSG", *COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Proceedings of the Conference (Université de Montrèal)* (San Francisco: Morgan Kaufmann) 1: 91–95.

6. Beck, Isabel L.; McKeown, Margaret G.; & Omanson, Richard C. (1987), "The Effects and Uses of Diverse Vocabulary Instructional Techniques", in Margaret G. McKeown & Mary E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 147–163.

7. Beck, Isabel L.; Perfetti, Charles A.; & McKeown, Margaret G. (1982), "Effects of Long-Term Vocabulary Instruction on Lexical Access and Reading Comprehension", *Journal of Educational Psychology* 74: 506–521.

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

9. Blachowicz, Camille, & Fisher, Peter (1996), *Teaching Vocabulary in All Classrooms* (Englewood Cliffs, NJ: Prentice-Hall/Merrill): 15–34.

10. Brown, Roger W. (1958), *Words and Things: An Introduction to Language* (Glencoe, IL: Free Press).

11. Budiu, Raluca, & Anderson, John R. (2001), "Word Learning in Context: Metaphors and Neologisms" *Technical Report CMU-CS-01-147* (Carnegie Mellon University, School of Computer Science).

12. Buikema, J., & Graves, M.F. (1993), "Teaching Students to Use Context Cues to Infer Word Meanings", *Journal of Reading* 36: 450–457.

13. Carbonell, Jaime G. (1979), "Towards a Self-Extending Parser", *Proceedings of 17th Annual Meeting of the Association for Computational Linguistics (University of California at San Diego)* (Morristown, NJ: Association for Computational Linguistics): 3–7.

14. Carr, E.M., & Mazur-Stewart, M. (1988), "The Effects of the Vocabulary Overview Guide on Vocabulary Comprehension and Retention", *Journal of Reading Behavior* 20: 43–62.

15. Carroll, J.B.; Davis, P.; & Richman, B. (1971), *The American Heritage Word Frequency Book* (New York: Houghton Mifflin).

16. Carter, D. (1989), "Lexical Acquisition in the Core Language Engine", *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics (University of Manchester)* (Morristown, NJ: Association for Computational Linguistics): 137–144.

17. Clarke, D.F., & Nation, I.S.P. (1980), "Guessing the Meanings of Words from Context: Strategy and Techniques", *System* 8: 211–220.

18. Copestake, Ann (1999), “The LKB System” [http://www-csli.stanford.edu/∼aac/lkb.html].

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

20. Dale, E. (1965), “Vocabulary Measurement: Techniques and Major Findings”, *Elementary English* 42: 895–901.

21. Deighton, L.C. (1959), *Vocabulary Development in the Classroom* (New York: Teachers College Press).

22. Ehrlich, Karen (1995), “Automatic Vocabulary Expansion through Narrative Context”, *SUNY Buffalo Computer Science Technical Report 95-09*.

23. 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.

24. Ericsson, K.A., & Simon, Herbert A. (1993), *Protocol Analysis: Verbal Reports as Data (Revised Edition)* (Cambridge, MA: MIT Press).

25. Fukkink, R.G., & de Glopper, K. (1998), “Effects of Instruction in Deriving Word Meaning from Context: A Meta-Analysis”, *Review of Educational Research* 68: 450–458.

26. Gambrell, L.B., & Almasi, J.F. (1993), “Fostering Comprehension through Discussion,” in L.M. Morrow, J.K. Smith, & L.C. Wilkinson (eds.), *Integrated Language Arts: Controversy to Consensus Reading* (Allyn & Bacon): 71–90.

27. Gipe, J.P. (1980), “Use of a Relevant Context Helps Kids Learn New Word Meanings”, *Reading Teacher* 33(4): 398–402.

28. Goodman, K.S. (1970), “Behind the Eye: What Happens in Reading”, in O.S. Niles & K.S. Goodman (eds.), *Reading Process and Program* (National Council of Teachers of English): 3–20.

29. Goerss, Betty L.; Beck, Isabel L.; & McKeown, Margaret G. (1999), “Increasing Remedial Students’ Ability to Derive Word Meanings from Context”, *Journal of Reading Psychology* 20: 151–175.

30. 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.

31. Guha, R.V. (1991), “Contexts: A Formalization and Some Applications” (Stanford University Ph.D. Thesis).

32. Hahn, Udo, & Schnattinger, Klemens (1998), “A Text Understander that Learns”, *COLING-ACL ’98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Proceedings of the Conference (Université de Montréal)* (San Francisco: Morgan Kaufmann) 1: 476–482.

33. Harmon, Janis M. (1999), “Initial Encounters with Unfamiliar Words in Independent Reading”, *Research in the Teaching of English* 33: 304–338.

34. Harmon, Janis M. (2000), “Assessing and Supporting Independent Word Learning Strategies of Middle School Students”, *Journal of Adolescent and Adult Literacy* 43: 518–527.

35. 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.

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

37. Hearst, Marti A. (1992), “Automatic Acquisition of Hyponyms from Large Text Corpora”, *Proceedings of the 15th International Conference on Computational Linguistics (COLING-92, Nantes, France* (Morristown, NJ: Association for Computational Linguistics), II: 539–545.

38. Ide, Nancy M., & Veronis, J. (eds.) (1998), Special Issue on Word Sense Disambiguation, *Computational Linguistics* 24(1).

39. Iwańska, Łucja M. & Zadrozny, Wlodek (1997), Special Issue on Context in Natural Language Processing, *Computational Intelligence* 13(3).

40. Jenkins, Joseph R.; Stein, Marcy L.; & Wysocki, Katherine (1984), “Learning Vocabulary through Reading”, *American Educational Research Journal* 21(4): 767–787.

41. Johnson-Laird, P.N. (1987), “The Mental Representation of the Meanings of Words”, *Cognition* 25: 189–211.

42. Kibby, Michael W. (1979), “Passage Readability Affects the Oral Reading Strategies of Disabled Readers”, *The Reading Teacher* 32: 390–396.

43. Kibby, Michael W. (1995), “The Organization and Teaching of Things and the Words that Signify Them”, *Journal of Adolescent and Adult Literacy* 39(3): 208–223.

44. Keirsey, 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.

45. Krasner, K.G. (1988), “A Study of the Effects of Instruction on Incidental Word Learning and on the Ability to Derive Word Meanings from Context”, Ph.D. dissertation (University of Delaware).

46. Kuhn, Melanie R., & Stahl, Steven A. (1998), “Teaching Children to Learn Word Meanings from Context: A Synthesis and Some Questions”, *Journal of Literacy Research* 30(1): 119–138.

47. Lenat, Doug (1998), “The Dimensions of Context-Space” [http://www.cyc.com/].

48. Levin, Beth (1993), *English Verb Classes and Alternations* (Chicago: University of Chicago Press).

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

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

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

52. McCarthy, John (1993), “Notes on Formalizing Context”, *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*.

53. Miller, George A. (1985), “Dictionaries of the Mind”, *Proceedings of 23rd Annual Meeting of the Association for Computational Linguistics* (Morristown, NJ: Association for Computational Linguistics): 305–314.

54. Miller, George A. (1991), *The Science of Words* (Scientific American Library).

55. Nagy, William E., & Anderson, Richard C. (1984), “How Many Words Are There in Printed School English?”, *Reading Research Quarterly* 19(3): 304–330.

56. Nagy, William E. & Herman, Patricia A. (1987), “Breadth and Depth of Vocabulary Knowledge: Implications for Acquisition and Instruction”, in Margaret G. McKeown & Mary E. Curtis (eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 19–35.

57. Nation, Paul, & Coady, James (1988), “Vocabulary and Reading”, in Ronald Carter & Michael McCarthy (eds.), *Vocabulary and Language Teaching* (London: Longman): 97–110.

58. Nist, S.L., & Kirby, K. (1986), “Teaching Comprehension and Study Strategies through Modeling and Thinking Aloud”, *Reading Research and Instruction* 25: 254–264.

59. Patberg, J.A.; Graves, M.F.; & Stibbe, M.A. (1982), "Vocabulary Instruction: Effects on Word Knowledge and Reading Comprehension", *Learning Disability Quarterly* 5: 202–215.

60. Pauling, Linus (1991/1994), Associated Press interview (1991), reprinted in his obituary, *Buffalo News* (21 Aug 1994): A9.

61. Perry, William G., Jr. (1970), *Forms of Intellectual and Ethical Development in the College Years* (New York: Holt, Rinehart, & Winston).

62. Pressley, Michael, & Afflerbach, P. (1995), *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading* (Lawrence Erlbaum Associates).

63. Pressley, Michael; Levin, Joel R.; & McDaniel, Mark A. (1987), "Remembering versus Inferring What a Word Means: Mnemonic and Contextual Approaches", in Margaret G. McKeown & Mary Curtis (eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 107–127.

64. Pustejovsky, James (1987), "On the Acquisition of Lexical Entries: The Perceptual Origin of Thematic Relations", *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (Stanford University)* (Morristown, NJ: Association for Computational Linguistics): 172–178.

65. Rapaport, William J. (2002), "A (Partial) Bibliography (in Chronological Order) of (Computational) Theories of Contextual Vocabulary Acquisition" [http://www.cse.buffalo.edu/∼rapaport/refs.vocab.html].

66. Rapaport, William J., & Ehrlich, Karen (2000), "A Computational Theory of Vocabulary Acquisition", in L. Iwańska & Stuart C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation* (Menlo Park, CA: AAAI Press): 347–375.

67. Rapaport, William J., & Kibby, Michael W. (2002), "Contextual Vocabulary Acquisition: A Computational Theory and Educational Curriculum", in Nagib Callaos, Ana Breda, & Ma. Yolanda Fernandez J. (eds.), *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002; Orlando, FL)* (Orlando: International Institute of Informatics and Systemics), Vol. II: Concepts and Applications of Systemics, Cybernetics, and Informatics I, pp. 261–266.

68. Reed, J.C. (1953), "Developing Word Recognition and Meaning Vocabulary in College", in Helen M. Robinson (ed.), *Corrective Reading in Classroom and Clinic* (Chicago: University of Chicago Press).

69. Schagrin, Morton L.; Rapaport, William J.; & Dipert, Randall R. (1985), *Logic: A Computer Approach* (New York: McGraw-Hill).

70. Schank, Roger C., & Rieger, Christopher J. (1974), "Inference and the Computer Understanding of Natural Language," *Artificial Intelligence* 5: 373–412.

71. Schatz, Elinore Kress, & Baldwin, R. Scott (1986), "Context Clues Are Unreliable Predictors of Word Meanings", *Reading Research Quarterly* 21(4): 439–453.

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

73. Shapiro, Stuart C. (1989), "The CASSIE Projects: An Approach to Natural Language Competence," in J.P. Martins & E.M. Morgado (eds.), *Proceedings of the 4th Portuguese Conference on Artificial Intelligence*, *Lecture Notes Artificial Intelligence* 390 (Springer-Verlag): 362–380.

74. Shapiro, Stuart C. (1992), "Artificial Intelligence", in Stuart C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence* (New York: John Wiley & Sons): 54–57.

75. Shapiro, Stuart C. (2000), "SNePS: A Logic for Natural Language Understanding and Commonsense Reasoning", in L. Iwańska & Stuart C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation* (Menlo Park, CA: AAAI Press): 175–195.

76. Shapiro, Stuart C., & Johnson, Frances L. (2000), "Automatic Belief Revision in SNePS", in C. Baral & M. Truszczynski (eds.), *Proceedings of the 8th International Workshop on Non-Monotonic Reasoning* .

77. Shapiro, Stuart C., & Rapaport, William J. (1987), "SNePS Considered as a Fully Intensional Propositional Semantic Network", in N. Cercone & G. McCalla (eds.), *The Knowledge Frontier* (New York: Springer-Verlag): 262–315.

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

79. Shapiro, Stuart C., & Rapaport, William J. (1992), "The SNePS Family", *Computers and Mathematics with Applications* 23: 243–275.

80. Shapiro, Stuart C., & Rapaport, William J. (1995), "An Introduction to a Computational Reader of Narratives", in J.F. Duchan et al. (eds.), *Deixis in Narrative* (Hillsdale, NJ: Lawrence Erlbaum Associates): 79–105.

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

82. Stalnaker, Robert C. (1999), *Context and Content* (New York: Oxford University Press).

83. 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.

84. Sternberg, Robert J.; Powell, Janet S.; & Kaye, Daniel B. (1983), "Teaching Vocabulary-Building Skills: A Contextual Approach", in A.C. Wilkinson (ed.), *Classroom Computers and Cognitive Science* (New York: Academic Press): 121–143.

85. Swanborn, M.S.L., & de Glopper, K. (1999), "Incidental Word Learning while Reading: A Meta-Analysis", *Review of Educational Research* 69: 261–285.

86. Thorndike, Edward L. (1917), "Reading as Reasoning: A Study of Mistakes in Paragraph Reading", *Journal of Educational Psychology* 8(6): 323–332.

87. van Daalen-Kapteijns, M.M., & Elshout-Mohr, M. (1981), "The Acquisition of Word Meanings as a Cognitive Learning Process", *Journal of Verbal Learning and Verbal Behavior* 20: 386–399.

88. van Daalen-Kapteijns, M.M.; Elshout-Mohr, M.; & de Glopper, K. (2001), "Deriving the Meaning of Unknown Words from Multiple Contexts", *Language Learning* 51: 145–181.

89. *Webster's New International Dictionary of the English Language* (Merriam, 1937).

90. Werner, Heinz, & Kaplan, Edith (1952), "The Acquisition of Word Meanings: A Developmental Study", *Monographs of the Society for Research Child Development*, 15 (1950) 51(1).

91. Wesche, Marjorie, & Paribakht, T. Sima (eds.) (1999), "Incidental L2 Vocabulary Acquisition: Theory, Current Research, and Instructional Implications", *Studies in Second Language Acquisition* 21(2).

92. Wilhelm, J.D. (2001), *Improving Comprehension with Think-Aloud Strategies* (New York: Scholastic Professional Books).

93. Wilks, Yorick, & Fass, Dan (1992), "The Preference Semantics Family," *Computers and Mathematics with Applications* 23: 205–221.

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