Contextual Vocabulary Acquisition: From Algorithm to Curriculum

William J. Rapaport[1] and Michael W. Kibby[2]

[1]Department of Computer Science and Engineering, Department of Philosophy, Department of Linguistics, and Center for Cognitive Science
rapaport@buffalo.edu

[2]Department of Learning and Instruction and Center for Literacy and Reading Instruction
mwkibby@gmail.com

http://www.cse.buffalo.edu/~rapaport/CVA/
State University of New York at Buffalo, Buffalo, NY 14260

Abstract

Deliberate contextual vocabulary acquisition (CVA) is a reader's ability to figure out *a* (not *the*) meaning for an unknown word from its "context", without external sources of help such as dictionaries or people. The appropriate context for such CVA is the "belief-revised integration" of the reader's prior knowledge with the reader's "internalization" of the text. We discuss unwarranted assumptions behind some classic objections to CVA, and present and defend a computational theory of CVA that we have adapted to a new classroom curriculum designed to help students use CVA to improve their reading comprehension.

Running head: Contextual Vocabulary Acquisition
Keywords:
Artificial Intelligence, Education, Linguistics, Philosophy; Instruction, Language acquisition, Language understanding, Learning, Reasoning, Semantics; Knowledge representation, Logic, Symbolic computational modeling.

0. From Guise Theory to Contextual Vocabulary Acquisition

0.1. An Opening Anecdote

When Hector-Neri Castañeda was ill, Randall R. Dipert and I (WJR) visited him in Bloomington. He was weak and tired, and his speech was even harder to decipher than it had been when he was healthy. He was happy to see us, however, and, after a few preliminaries, suggested that we have dinner later at his favorite macrobiotic restaurant. The restaurant was exceedingly noisy, rendering dinner conversation virtually impossible. Although he was tired after dinner, he insisted that we come back to his house to talk philosophy. And, as soon as we began our philosophical conversation, his stamina returned, his voice became stronger and clearer, and we were back with the Hector of old. Philosophy gave him strength.

0.2. Castañeda's Guise Theory.

Guise theory (Castañeda 1972, 1999, and many works in between; for a summary, see Rapaport 2005b) has two goals: to serve as a theory of the mechanism of reference and to serve as a theory of the world as it appears to us. And it has two principal sources: Frege's (1892) puzzle of reference and the "univocality" of language. Frege's puzzle concerns the way in which the President of the US is the "same" as the Commander-in-Chief of the US Armed Forces. The univocality of language is the claim that talk and thought about truth and reality behaves exactly the same as talk and thought about falsehood and fiction: Language treats objects of both kinds of talk and thought in the same way. This is associated with a principle of "metaphysical internalism": Talk and thought about the world is internal to experience; the (experienced) world must be understood from the "inside", not "externally"—from a first-person point of view, not from God's point of view or Nagel's (1986) "view from nowhere". Metaphysical internalism has proven useful in artificial intelligence (AI) and cognitive modeling (Shapiro & Rapaport 1987, 1991).

According to guise theory, there are properties, sets of properties (called "guise cores"), and an "individuating operator" (c) that maps guise cores to guises. E.g., where Round is the property of being round and Square is the property of being square, c{Round, Square} is the guise: the round square. Guises are, roughly, things under a description, or facets of objects, or roles played by objects, or intentional objects of thought.

There are also two modes of predication: internal and external. Internal predication is membership in a guise core: c{…F…} is (internally) F; the round square is both round and square, because c{Round, Square} is (internally) Round and Square (because both Round and Square ∈ {Round, Square}.

There are several kinds of external predication: Consubstantiation (C*) is an equivalence relation among actual objects: The morning star is consubstantiated with the evening star: C*(c{Morning, Star}, c{Evening, Star}). One way of saying that the morning star (externally) has a certain property is as follows: Let a = c{…F…} be a guise; let a[G] $=_{def}$ c({…F…} ∪ G); then: a is (externally) G if and only if C*(a, a[G]). E.g., the morning star is a planet, because C*(c{Morning, Star}, c{Morning, Star, Planet}). Consubstantiation can help solve Frege's puzzle as follows: Let a, b be guises; then a is the same as b if and only if C*(a, b). E.g., the morning star is the evening star, because C*(c{Morning, Star}, c{Evening, Star}). And a "really" exists if and only if $\exists x$C*(a, x). Finally, there is a semi-lattice (i.e., a partially ordered set with least upper bounds) of consubstantiated guises; real objects are infinitely propertied, multi-faceted, "Leibnizian" individuals at the "apex" of such a semi-lattice.

Other kinds of external predication include: consociation (C**), an equivalence relation among guises that a mind has "put together" in a "belief space"; transubstantiation, which is identity across time; and transconsociation, which is identity across works of fiction.

The Meinongian theory developed in Rapaport 1978 and elsewhere presented an ontology for epistemology, or what Castañeda called a "phenomenological ontology": an ontology of the first-person, mental objects of thought. A theory of univocal reference presented in Rapaport 1981: Whereas formal languages typically have a *total* semantic interpretation function, natural languages only have a *partial* semantic interpretation function. To give a semantics for non-referring noun phrases, one must either *reform the syntax* (as Russell and Quine suggested) or *expand the semantics*, in order to turn the partial function into a total function. The latter can be accomplished by using Meinongian objects. This was later applied to give a semantics for the SNePS semantic network, knowledge-representation and reasoning system (Shapiro & Rapaport 1987; see §2.4.4, below). Rapaport 1981 suggested a way to extend the theory: Instead of viewing Meinongian objects as structures of *properties*, where the meaning of a term such as 'bachelor' (using the double-bracket notation for the meaning-function, this will be denoted as: [['bachelor']]) might be the Meinongian object <Male, Marriage-eligible> (Rapaport's analogue of the guise c{Male,

Marriage-eligible}),[1] we can view them as structures of *predicates*, or *contexts*, or "*co-texts*" (see §3.4, below): [['bachelor']] = < '…is male', '…is marriage-eligible'>. (Compare latent semantic analysis, which considers meaning "as [a] decontextualized summary of all experiences with a word".) More generally, let *w* be a word (or other linguistic expression), let *S* be a speaker, and let *t* be a time; then we can let $[[w]]_{S,t}$ = < C(*w*) : C(*w*) is a co-text (or context, or predicate) of *w* heard by *S* before *t* >. I.e., a meaning for a word *w* for a speaker *S* at time *t* is the set of all contexts (or co-texts, or predicates) containing *w* that *S* has heard up till *t*.

This suggests an interesting research project in computational contextual vocabulary acquisition": How can we compute $[[w]]_{S,t}$? One way is to consider the semantic network $N_{S,t}$ representing the propositions that *S* believes at *t*. We can describe $N_{S,t}$ from *w*'s point of view if and only if we can figure out (i.e., compute) *S*'s meaning at *t* for unknown word *w* from context. And this has led us from Castañeda's guise theory to…

1. A Computational Theory of Contextual Vocabulary Acquisition

Readers often don't—or don't want to—look up unfamiliar words "in the dictionary". Sometimes, they don't realize that a word is unfamiliar, reading past it as if it weren't there: Even "skilled readers" skip "about one-third of the words" (Brysbaert et al. 2005), and words that are highly likely to be chosen in a cloze task are skipped more than others (Rayner & Well 1996); paradoxically, "children may misread or ignore unfamiliar words without jeopardizing comprehension" (Bowey & Muller 2005, citing Share 1999). Other times, readers realize that they don't understand a word but are too lazy (or embarrassed) to do anything about it: Perhaps—discouraged by previous, unsuccessful attempts to look words up—they hope that the word isn't important and won't be used again.

Or they might be curious as to what the unfamiliar word might mean, i.e., what the author had in mind when using it. If no one is around to ask, or if the only people around don't know the word, then the reader can look it up in "the" dictionary. But what if no dictionary is handy? Or the only one doesn't contain the word? Or it does, but the definition seems inappropriate for the context, or is unhelpful? With the exception of learner's dictionaries designed primarily for ESL readers, dictionaries are often difficult to use and their definitions difficult to interpret (Miller 1985, 1986; Miller & Gildea 1985; Rapaport & Kibby 2007).

Alternatively, readers can "figure it out" from the "context". We call this "contextual vocabulary acquisition" (CVA). Just as giving a person a fish feeds them for a day, but teaching them to fish feeds them for a lifetime, so

[1] On "marriage-eligible", see Cole 1999.

giving a reader a definition tells them one word's meaning, but *teaching* them CVA enables them to become better readers. We make four principal claims: (1) If "[t]exts ... are ... the soil in which word-meaning understandings are grown" (Wieland 2008), then the reader's prior knowledge is the water enabling that growth. (2) A procedure for successful CVA can be expressed in terms so precise that it can be programmed into a computer and (3) taught to human readers, (4) helping them improve both vocabulary and reading comprehension.

How does one do CVA? What is the "context" for acquiring new vocabulary? Readers do CVA "incidentally" (§2.3). They can also do it "deliberately" (§2.4; cf. Hulstijn 2003). And they can be taught how to do it. But how well are they taught it? What *should* we teach readers to do when confronted with an unfamiliar word? And can it be taught in a way that readers can use in any situation?

In a future "golden age" of cognitive science and AI, computers might be able to fully converse with humans. Although no computer programmed to process natural language can yet convince a native speaker that it "understands" the language (Turing 1950, Shieber 2004, Rapaport 2006a), there are many programs that process natural-language text—that can parse it, construct semantic interpretations of it, answer simple questions about it, or do information retrieval (Jurafsky & Martin 2008, Gunning et al. 2010). Such a computer confronted with an unfamiliar word should be able to do (or fail to do) anything that a human could: It could attempt to look the word up in an online lexical resource. But if the word is new or not in the relevant database, then it might have to "figure out"—i.e., compute—a meaning for the unfamiliar word from "context". Can a computer do this? Under certain (reasonable) circumstances, it can.

Moreover, we can adapt its methods for doing this to teach human readers to compute meanings in the same way. This reverses the usual way in which computers are programmed (or "taught") to do certain cognitive tasks. In "good old-fashioned", symbolic AI (Haugeland 1985), if we know how to explicitly teach a cognitive task to humans (e.g., how to play chess, solve calculus problems, prove logic theorems), then we can explicitly program a computer to do that task in pretty much the same way that humans do it. If we don't know how to teach some task—how would you *teach* a human to see?—then it is difficult (though not necessarily impossible) to program a computer to perform that task. Fortunately, meaning-vocabulary acquisition is the first kind of task.

Our research group (including reading educators, AI researchers, and a philosopher of language and logic) has been teaching a computer to figure out meanings of unfamiliar words from context, in order to see if what we learn in teaching *it* can help us teach *students* better. This is a two-way street: We are also improving our program based

on observations of good readers doing CVA (Wieland 2008). But our study of the CVA literature in computational linguistics, psychology, first- and second-language (L1, L2) acquisition, and reading education shows that each discipline tends to ignore the others' literatures (Rapaport & Ehrlich 2000, Wieland 2008, Rapaport 2010). Thus, our project has twin goals: (1) to develop a computational, truly interdisciplinary, cognitive theory of deliberate CVA, and (2) to adapt our computational CVA strategies to an educational curriculum for teaching them to students.

2. The Nature of Contextual Vocabulary Acquisition

2.1 What Is Word Meaning? For many reading educators, 'sense' is just a synonym for 'meaning'; for philosophers, 'sense' is a technical term. Frege (1892) analyzed meaning into two components: '*Sinn*' and '*Bedeutung*'. Although both of these can be translated as "meaning", the former is usually translated 'sense' in English, the latter as 'reference', 'referent', or 'denotation'. For Frege, a *Bedeutung* was something in the world that a linguistic expression referred to; e.g., 'snow' refers to the cold, wet, white stuff that precipitates from the sky during winter—the actual stuff, not that description of it. However, 'unicorn' has no referent, because there are no unicorns. There are pictures and stories about them, but (unfortunately) they don't exist. But 'unicorn' is far from meaningless: It has a sense (a *Sinn*), even though there is no referent. But 'snow' has a referent *in addition to, and determined by*, its sense.[2] You can identify real snow if you know that the sense of 'snow' is (roughly) *cold, wet, white stuff that precipitates from the sky during winter*. And you could (probably)[3] identify a real unicorn if you ever saw one (even though you won't). So, all words have senses, but not all words have referents. Fregean senses are "objective"; each unambiguous word has just one sense, which, somehow, all minds "grasp".

'Meaning' might be a more neutral term. But we should not talk about "the" meaning of a word. A word has *many* meanings. Not only are most words ambiguous (polysemous), but each of us has our "own", psychologically unique meaning for each word. Frege considered these unscientific, wanting to de-psychologize meaning. The meanings computed by our CVA programs *are* "psychological" in this way, hence similar to, but not exactly the same as, Fregean senses. They are closer to being "Meinongian objects" (Meinong 1904, Castañeda 1972, Parsons 1980, Rapaport 1981, Zalta 1983) and are the focus of the semantic theories of "cognitive linguists" (Langacker

[2]Church 1956. Strictly speaking, for Frege, every linguistic expression has a "sense" (*Sinn*), and some, but not all, *senses* (*Sinne*) have "referents" (*Bedeutungen*). E.g., both 'snow' and 'unicorn' have senses, the *sense* of 'snow' has a *referent* (the cold, wet, white stuff), but the *sense* of 'unicorn' doesn't. Multiple "senses" (*Sinne*) can have the same "referent" (*Bedeutung*); e.g., the senses of 'the morning star' and 'the evening star' both refer to Venus.

[3]It has been said that if we saw a creature satisfying the definition of 'unicorn', we might refuse to call it that, because we know that unicorns don't exist. So, if we saw such a creature, we might call it "unicorn-*like*" but wouldn't say that it *was* the mythical creature (see Associated Press 2008).

1999, Talmy 2000, Jackendoff 2002). Problems about how we can understand each other if we all mean something different by our words can be overcome (see §4.1.3 and Rapaport 2003a). So, we use the indefinite noun-phrase '*a* meaning *for* a word' (rather than the more common, definite noun-phrase '*the* meaning *of* a word') to emphasize that meanings produced by CVA are *hypotheses* constructed and assigned to words by the reader, rather than "correct" (dictionary-style) definitions that "belong" to words:

> [A] word hasn't got a meaning given to it, as it were, by a power independent of us, so that there could be a kind of scientific investigation into what the word *really* means. A word has the meaning someone has given to it. (Wittgenstein 1958: 28.)

One must be careful to steer clear of Humpty Dumpty's claim that a word "means just what I choose it to mean—neither more nor less" (Carroll 1896), as Wittgenstein's last sentence suggests. There are similar ideas in contemporary cognitive science: Lakoff & Johnson (1980) decry the implication from the usual terminology that "words ... have meanings in themselves, independent of any context or speaker" (p. 459). Clancey (2008) says, "We cannot locate meaning in the text ...; [locating meaning is an] active, dynamic process..., existing only in interactive behaviors of cultural, social, biological, and physical environment-systems." And Elman (2007) says, "Following … Rumelhart …, I will propose that rather than thinking of words as static representations that are subject to mental processing—*operands*, in other words—they might be better understood as *operators*, entities that operate directly on mental states in what can be formally understood as a dynamical system." "[W]ords should be thought of not as having intrinsic meaning, but as providing cues to meaning" (Elman 2009).[4]

2.2. What Is CVA? By 'CVA', we mean the acquisition by a reader of a meaning for a word in a text by means of reasoning from textual clues and prior knowledge (including language knowledge and hypotheses developed from prior encounters with the word), but without external sources of help such as dictionaries or people. Although CVA can be used in conversation, watching TV, etc., we focus on CVA during reading. Everything we say should carry over to spoken domains (Gildea et al. 1990, Beals 1997, Aist 2000). However, Cunningham & Stanovich 1998 provide evidence that "conversation is not a substitute for reading" in terms of the benefits of reading for improving not only vocabulary but general intelligence (as well as evidence that watching TV has *negative* effects!).

---

[4]Many authors write of "cues"; others, of "clues"; some (Beck et al. 1983), of both. 'Cue' suggests a textual element that prompts the reader, perhaps unconsciously, to think of something. 'Clue' suggests a textual element that a knowledgeable reader can use (perhaps consciously) to infer something. Thus, not all cues are clues, nor vice versa. The terms seem interchangeable in the literature, but we will use them as mentioned here, except when quoting.

For an example of CVA as we are studying it, consider this passage:

> Trains go almost everywhere, and tickets cost roughly two dollars an hour for first-class travel (first-class Romanian-style that is, with tatterdemalion but comfortably upholstered compartments ....) (Tayler 1997.)

One informant thought out loud as follows about *tatterdemalion*'s meaning:

> It kind of makes me feel like they're not chic or really nice but they are comfortable .... And it is first class, so it seems as if they may be worn [out] .... It's "Romanian-style"; they talk about that it's a pretty poor country. (Cited in Wieland 2008.)

Although this reader did not offer a definition at this point in the protocol, she appears to understand the word as connoting something roughly second-class. Our computational system reasoned as follows (Schwartzmyer 2004): [5]

> Comfortable and tatterdemalion are related by "but". If *x* and *y* are related by "but", and *y* is a positive attribute, then *x* is a negative attribute, equivalent to a negative quality. Comfortable is a positive attribute; so, being tatterdemalion is a negative attribute. Romania is poor, so its trains are poor. Tickets for first-class travel are expensive. First-class travel is comfortable and of high quality. If tickets for travel cost $2, then they are not expensive, so this is not first-class travel. If something is not first-class travel, and trains (used for this travel) have parts whose properties are described in terms of the above "but" schema, then one of these properties will be equivalent to being second rate. So "first-class Romanian-style" travel is really second-rate travel. Therefore, this train is really second-class travel, so being tatterdemalion is being second rate.

CVA discovers neither a word's "correct" meaning (whatever that might be) nor (necessarily) the author's intended meaning. Rather, it is a process of (a) developing *a* theory about *a* meaning that a particular use of a word in some particular textual passage *might* have, (b) *temporarily* assigning that meaning to the word, and (c) testing that hypothesis on future encounters with the word. The reader only has to determine *a* meaning for the word (as it appears in the text) *that enables understanding the text sufficiently to continue reading*. Following Lakoff & Johnson, our claim is that a meaning for a word depends on both its context and the reader.

Most CVA researchers believe it possible to "figure out" a meaning for a word "from context". Other terms for this include 'construct', 'deduce', 'derive', 'educe', 'guess', 'infer', and 'predict'. Because CVA is computable (as evidenced by the existence of our algorithms), one of us (WJR) prefers the phrase "*compute* a meaning". That is what our software does, and what our algorithm-based curriculum teaches. It is also what—on the computational

[5]This is a translation of the computer transcript into full English sentences.

theory of mind—human readers do (but see Rapaport 1998). 'Deduce' or 'infer' are too narrow; Simon (1996) observes that it is "more accurate to say that" a text "suggests" meanings than that a reader "infers" meanings from it, but perhaps these are two sides of one coin. We dislike "guess", connoting randomness, especially for its lack of guidance for readers doing CVA (see §§2.4.2, 7.4, below). However, the phrase "the reader computes a meaning" is awkward at best, so we shall use "figure out". (In any case, "figure out" may be metaphorical for "compute"!)

2.3 Incidental CVA. CVA is not restricted to fluent readers faced with a new word, incrementally increasing their vocabulary. Most of our vocabulary is acquired this way, in a bootstrapping process: 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 (Nagy & Anderson 1984, Nagy & Herman 1987). The *average* number of word families (e.g., 'help', 'helps', 'helped', 'helping', 'helper', 'helpless', 'helpful' are one word-family) known by high-school graduates has been estimated at between 45,000 (Nagy & Anderson 1984) and 60,000 (Miller 1991). Students who read a lot may know twice that (Miller 1991). But learning even 45,000 words by age 18 means learning on average 2500 words/year; yet no more than 400 words/year are directly taught by teachers (Nagy & Anderson 1984)—4800 words in 12 years of school. Therefore, some 90% of the words we know and understand must have been learned from oral or written context. CVA is not a once-in-a-while thing; it conservatively averages almost 8 words learned per day (Nagy & Anderson 1984). Most of this vocabulary acquisition is "incidental" (Nagy et al. 1985; Christ 2007 discusses incidental CVA from our perspective): It is very likely the result of unconscious, inductive inference (Reber 1989, Seger 1994).

2.4 Deliberate CVA. Some CVA is the result of "deliberate" (conscious, active) processes of hypothesizing meanings for unfamiliar words from context. The psychology, L1, and L2 literatures advise us to look for definitions, synonyms, antonyms, examples, apposition, comparison, contrast, etc. (Ames 1966, Clarke & Nation 1980, Van Daalen-Kapteijns & Elshout-Mohr 1981, Sternberg et al. 1983, Sternberg 1987, Kibby 1995, Blachowicz & Fisher 1996, Wesche & Paribakht 1999, Baumann et al. 2003). But these are rare, and, once found, what should we do with this information? Deliberate CVA instruction in classroom settings has not fared well:

2.4.1 Mueser & Mueser. Mueser & Mueser (1984; Mueser 1984) have a CVA workbook (once used by a nationwide tutoring center) that begins with a multiple-choice pre-test on the definitions of a set of words, followed by a set of 4- or 5-sentence paragraphs using each word "in context", followed by the same multiple-choice quiz as a post-test. But one sentence of each paragraph contains a definition of the word! This is overly helpful.

2.4.2 Nation et al. Clarke & Nation (1980) offer the following strategy, which begins helpfully:

> **step 1:** "look at the word … and its surroundings to decide on the part of speech." This depends only on the reader's knowledge of grammar (which is "in" the reader's mind, not in the text; see §3, below).
>
> **step 2:** "look at the immediate grammar context of the word, usually within a clause or sentence." This presumably gives such information as "what is done to what, by whom, to whom, where, by what instrument, and in what order" (Bruner 1978).
>
> **step 3:** "look at the wider context of the word usually beyond the level of the clause and often over several sentences", presumably looking for information of the sort recommended by those cited above (§2.4).
>
> **step 4:** "... *guess...the word* and check...that the guess is correct" (our emphasis).

This is more useful than Mueser's technique. But how should the poor reader combine the data gathered in steps 1–3 to produce a "guess"? And if the reader was supposed to guess in the first place, why gather data? In particular, all the important work that we are concerned with is hidden in the first conjunct of step 4, reminiscent of a famous Sidney Harris cartoon[6] showing a complicated mathematical formula, in the middle of which appears the phrase, "then a miracle occurs", about which an observer comments, "I think you should be more explicit here ...".

To be fair, Clark and Nation point out that 'guess' really means "infer" (p. 211n.1), but they don't offer the reader any guidelines for *how* to make the inference.[7] Later on, they say:

> Let us now look at each step *in detail*. [emphasis added]... 4. After the learner has gone through the three previous steps of part of speech, immediate context, and wider context, he should attempt to guess the meaning and then check his guess. There are three ways of checking ....

This is supposed to provide "detail". Is something missing between "guess the meaning and then check his guess" and "There are three ways of checking"? Nation & Coady (1988; Nation 2001) offer a *five*-step "elaboration of" Clarke & Nation's strategy: Steps 1–3 are similar—here are steps 4 and 5:

---

[6][http://www.sciencecartoonsplus.com/gallery.htm]

[7] Mondria & Wit-de Boer (1991) say explicitly that to guess is to infer from context. And Loui (2000) says:

> One of my favorite questions is, "what information is missing?" By looking for additional relevant information, the student must literally see what is not there: the student must exercise the imagination. Sometimes, when there is insufficient information, the student must make an imaginative educated guess. Guessing is an important skill in many disciplines. In computer programming, guessing is called "debugging". In economics and meteorology, guessing is called "forecasting". *In science, guessing is called "hypothesis formation."* In medicine, guessing is called "diagnosis". [our emphasis]

4. Guessing the meaning of the unknown word.

5. Checking that the guess is correct.

But splitting one conjoined step into two single steps is not a useful elaboration.

What's missing is precisely the level of detail that must be provided to a computer to enable it to figure out a meaning. You cannot ask a computer to guess; you *must* "be more explicit". But *if we can tell a computer what to do in order to "guess", then we should also be able to tell a student* (Knuth 1974, Schagrin et al. 1985: xiii). We do this in our curriculum (§7).

2.4.3 Sternberg et al. Sternberg (1987; Sternberg & Powell 1983; Sternberg et al. 1983) calls CVA "learning from context". But there are two different notions: (1) CVA is, roughly, figuring out a meaning for a word from its textual context plus the reader's background information (we will be more precise about this in §3). Authors assume that readers already know a meaning for each word in the text, so do not necessarily construct the text to purposefully provide information for learning a meaning for an unfamiliar word in it. This must be distinguished from (2) learning word meanings from such purposeful contexts: teaching a meaning for a word by presenting the word in a (sentential) context constructed so that the reader who may not know the intended meaning is able to learn it from the context. Sternberg et al. 1983 use "the example, *carlin*, which means *old woman*" (p. 125), presenting its meaning in this specifically-designed context: "Now that she was in her 90s, the once-young woman had become a carlin." A sentence constructed to provide a meaning-rich context for an unknown word should make it easy for the reader to figure out a meaning for the word. But not all contexts are created for such purposes. CVA enables readers to figure out meanings from *any* context.

Sternberg et al. 1983 (cf. Sternberg 1987) contrasts three theories of vocabulary "building": rote learning, keyword, and "learning from context". Rote learning is simply memorizing a word and its definition, both of which are given to the student. The keyword method *also* requires that the definition be *given* to the student; it then requires students to come up with imagistic links to improve their ability to remember the definition; e.g., for 'carlin', Sternberg suggests a mental image of an old woman driving a car. The proper contrast is *not* between these and CVA, but between these and learning meanings from "pedagogical" contexts. Most of the experiments that Sternberg cites probably considered learning from contexts *designed* for teaching, not from *naturally-occurring* contexts, as with CVA. The rest of Sternberg's paper *is* about CVA as we do it, so our comments might be moot,

though they do suggest that perhaps many others who seem to be opposed to CVA are really opposed to learning from contexts designed for teaching. We return to this in §4.

However, definition-providing contexts are the main ones that Sternberg uses as examples. One of them deserves some discussion for another reason:

> Although for the others the party was a splendid success, the couple there on the blind date was not enjoying the festivities in the least. An *acapnotic*, he disliked her smoking; and when he removed his hat, she, who preferred "ageless" men, eyed his increasing *phalacrosis* and grimaced. (Sternberg 1987.)

This passage contains *two* unknown words, both of whose meanings are easily figured out. Sternberg et al. (1983; cf. Sternberg 1987) correctly cites "Density of unknown words" as negatively affecting CVA. But this is not an example of the density problem, because the last sentence is easily rewritten to separate the two unknowns, and neither is needed in order to figure out the meaning of the other. Everyone we have shown this passage to figures out that 'phalacrosis' means either "baldness" or "grey hair", reasoning roughly as follows: Because she grimaced when she saw his phalacrosis, she doesn't like it. Because she likes ageless men, this man is not ageless. Hence, phalacrosis is a sign of male aging. Therefore, because the phalacrosis became visible when the hat was removed, it's either baldness or grey hair (more likely the former, because the latter is not unique to males). It does, in fact, mean "baldness".[8] Has a reader who decided it meant "grey hair" figured out an "incorrect" meaning? Technically, yes; but does it matter? Such readers would have figured out *a* (reasonable) meaning enabling them to understand the passage and continue reading. Admittedly, the reader might miss the author's intended meaning, but if the difference between baldness and grey hair becomes important later, the reader should be able to revise the hypothesis at that time. And if it does not become important, then missing the intended meaning is unlikely to be important now.

Sternberg shares our goals: to produce a theory of how to do and teach CVA. But his theory (Sternberg et al. 1983) is vague: His eight cues and seven mediating factors are devoid of detail, containing no instructions on what to do with them (except for an example or two; background knowledge plays a much larger role in our theory than in his). Sternberg et al.'s (1988) "general strategy for context use" (i.e., for CVA) compounds this:

> **step 1:** "Attempt to infer the meaning of the unknown word from the general context preceding the word ...."
>
> **step 2:** "Attempt to infer the meaning … from the general context that follows the word ....";

[8]In 2003, this was hard to find in any dictionary. By 2007, Google returned several websites with definitions, the most curious of which was *Wikipedia* (2007), which (erroneously) called it a "nonce" word, citing one of WJR's websites about precisely this sentence [http://www.cse.buffalo.edu/~rapaport/CVA/acapnotic.html]!

step 3: "Attempt to infer the meaning … by looking at the word parts ..." (i.e., its morphology);[9]

step 4: "If it is necessary ["to understand the word's meaning in order to understand the passage ... in which it is used"], estimate how definite a definition is required; if it is not necessary, further attempts to define the word are optional ...."

step 5: "Attempt to infer the meaning … by looking for specific cues in the surrounding context ...."

step 6: "Attempt to construct a coherent definition, using internal and external cues, as well as the general ideas expressed by the passage and general world knowledge ...."

step 7: "Check definition to see if meaning is appropriate for each appearance of the word in the context ...."

This *appears* to be more detailed than Clarke & Nation's strategy. However, steps 1 and 2 do not specify *how* to make the inference from context, nor how to relate these *two* inferences. Step 5 appears at best to be part of the needed specifications for steps 1 and 2, and at worst to merely repeat them. In any case, step 5 is no more detailed than Clarke & Nation's steps 3 or 4. Questions remain: *What* should the reader do with the information found in steps 1–5? *How* should the reader make the required inference? And *how* should the reader "construct" a definition (step 6)? Although Sternberg (1987) goes into considerably more detail, he offers a grab-bag of techniques, not an algorithm that could be systematically applied by a reader.

2.4.4 The Computational Approach. Although many authors suggest which contextual clues to look for, few (if any) say what to *do* with them once found. How should global and local text comprehension be employed? What reasoning and other cognitive processes are useful? And how should prior knowledge be applied?

Previous views of CVA privilege specific textual clues (§2.4). But "the reality may be that instruction in morphemic and contextual analysis—particularly when implemented in more naturalistic experimental settings—is simply too far removed from text comprehension to influence students' understanding directly" (Baumann et al. 2003). We privilege the reader's comprehension, prior knowledge, and reasoning, which bring at least as much

[9]Sternberg calls the use of morphological information "internal" context (as opposed to "external" co-text). However, a reader's ability to use morphological or etymological information depends entirely on the reader's prior knowledge (e.g., of the usual meanings of affixes). As Sternberg notes, internal context can't be used in isolation from external context (Sternberg & Powell 1983, Sternberg et al. 1983). E.g., 'inflammable' looks like it ought to mean "not flammable", because of the (apparently negative) prefix 'in-'. However, it is really synonymous with 'flammable', which might be determined from external context. Consequently, there is really nothing especially "contextual" (in the narrow, ("co-")textual sense) about the use of morphology. To add this to our computational system, our algorithm would simply need a procedure that checks for morphological information (gathered from the grammatical parse of the sentence containing the unknown word) and then uses prior knowledge to propose a meaning, which would then have to be checked against "external" contextual clues. On the other hand,

power and information to CVA as the text provides in terms of explicit clues: Readers' understanding is a function of how much *textual analysis* (of the sort involved in our CVA procedures) is involved, *not* a function of merely *looking* for clues that are either obvious or possibly non-existent. One other strategy is *slightly* more explicit: Buikema & Graves (1993) have the students *brainstorm* about what the word might mean, based on textual cues and past experience. This is fair, but not very precise; nor is it easily done, easily taught, or easily computed.

Indeed, the ability to "compute" a meaning is crucial, insofar as computer science is best understood as the natural science of procedures (rather than as a study of machines; Shapiro 2001, Denning 2007). Readers need to be taught a procedure that they can easily follow and that is almost guaranteed to enable them to figure out a meaning for a word from context. Unfortunately, little (if any) of the *computational* research on the *formal* notion of contextual reasoning is directly relevant to CVA (Guha 1991, McCarthy 1993, Iwańska & Zadrozny 1997, Lenat 1998, Stalnaker 1999—Hirst 2000 suggests why; see §3.2, below). Knowing more about the nature of context, having a more precise theory of CVA, and knowing how to teach it will allow us to more effectively help students identify context clues and know better how to use them, leading to improved reading comprehension.

Learning concepts and their words—especially when the concept is not part of the reader's prior knowledge, more especially when the reader has the prior knowledge needed to learn it quickly[10]—increases the reader's conception of the world and helps "students expand their ability to perceive similarities, differences, and order within the world" (Kibby 1995). 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 "assimilating" (Piaget 1952), "integrating" (Kintsch 1988), "consolidating" (Hansson 1999), or "connecting" (Storkel 2009) the newly learned concept with prior knowledge, which might include inference, belief revision, or reorganizing existing cognitive schemata.

Spelling out all the steps of inference, as we sketched for 'tatterdemalion' and 'phalacrosis', constitutes an *algorithm* for CVA. The best way to express such an algorithm is in a computer program, which has the extra benefit that it can be easily tested by implementing and then executing the program to see what it does. This is what

---

it would be interesting to have readers use CVA to learn the meanings of *affixes* from context! (We do incorporate "internal context" into our curriculum; see §7.3, below.)

[10]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, but even an unsophisticated reader has the prior knowledge necessary to learn this concept. By contrast, 'kurtosis' refers to the relative flatness or peakedness of a frequency distribution as contrasted with a normal distribution. Not only would readers not know this word, but they would have to know statistics to learn it (Kibby 1995).

cognitive scientists mean when they say that such computer programs *are* theories of psychological behavior (Simon & Newell 1962; but cf. Thagard 1984). We are investigating ways to facilitate readers' natural CVA by (1) developing a rigorous, computational theory of how context is used and (2) creating a systematic, viable curriculum for teaching CVA strategies, based on our AI algorithms and on analysis of CVA processes used by good readers.

Our computational theory of CVA was implemented (in Ehrlich 1995) in a propositional semantic-network knowledge-representation-and-reasoning system (SNePS; Shapiro & Rapaport 1987, 1992, 1995; Shapiro 2008). Other computational CVA systems have been developed by Granger 1977, Haas & Hendrix 1983, Berwick 1983, Zernik & Dyer 1987, Hastings & Lytinen 1994ab, Siskind 1996, et al. (see Rapaport & Ehrlich 2000).

Our computational system has a stored knowledge base containing SNePS representations of relevant prior knowledge. It takes as input SNePS representations of a passage containing an unfamiliar word. Processing begins with inferences drawn from these two, integrated information-sources.[11] When asked to define the word, it applies noun- and verb-definition algorithms (adjectives and adverbs are under investigation) that deductively search the integrated network for information of the sort that might be found in a dictionary definition, outputting a definition "frame" (Minsky 1974) or "schema" (Rumelhart 1980) whose slots are the kinds of features that a definition might contain (e.g., class membership, properties, actions, spatio-temporal information, etc.) and whose slot-fillers contain information gleaned from the network. (See §6; details of the underlying theory, representations, processing, inferences, belief revision, and definition algorithms are presented in Ehrlich 1995, 2004; Ehrlich & Rapaport 1997, 2005; Rapaport & Ehrlich 2000; Rapaport & Kibby 2007; Rapaport 2003b, 2005; Shapiro et al. 2007.)

In order to teach such an algorithm (minus implementation-dependent details) to students, it must be converted to a curriculum. We describe this in §7, below. But first we need to examine the notion of "context".

3. The Problem of "Context"

3.1 Word and Context. Probably no two researchers mean the same thing by 'context'. But we need to work with a reasonably precise characterization. The smallest useful *textual* context of a word would probably be a phrase containing the word as a grammatical constituent, typically the *sentence* containing it.

Which comes first: sentence meaning, or word meaning? It may seem obvious that (1) *word meanings are primary and sentence meanings depend on the meanings of the constituent words*; this underlies the standard

[11]This is equivalent to the techniques developed independently by Kintsch & van Dijk 1978 and in Discourse Representation Theory (Kamp & Reyle 1993). See Shapiro 1979, 1982; Rapaport & Shapiro 1984; Rapaport 1986b; Shapiro & Rapaport 1987, 1995.

compositional (or "recursive") view of semantics espoused by most contemporary linguists (Szabó 2007), as well as most approaches to vocabulary instruction and the use of dictionaries. There are, however, some clear exceptions, such as the lexicographer's method of determining word meanings from actual uses of the word (Murray 1977): "Lexicographers have to define words *in situ*, not in the abstract, removed from context" (Gilman 1989). Consequently, some researchers hold that (2) *sentence meanings are primary and word meanings depend on the meanings of their containing sentences*. Although Frege is well known for espousing (1), at one time he also held (2): "Only in the context [*Zusammenhange*] of a sentence [*Satz*] do words mean [*bedeuten*] something" (Frege 1884, §62; cf. §60).[12] Russell's (1905) theory of definite descriptions can be understood this way, too: One can't say what 'the' means, only what a sentence containing 'the' means—e.g., 'The present King of France is bald' means "there is one and only one present King of France and he is bald". No identifiable part of that meaning *is* a meaning for 'the'. And Quine (1961) noted that "the primary vehicle of meaning came to be seen no longer in the term but in the statement."[13] This has also been recognized in the field of vocabulary acquisition in the teaching of reading: "while context always *determines* the meaning of a word, it does not necessarily *reveal* that meaning" (Deighton 1959).

We maintain that *both* (1) *and* (2) are the case. Rather than try fruitlessly to determine an answer to this chicken-or-egg question, we take a holistic view of the situation (Saussure 1959; cf. Rapaport 2002, 2003a): Each individual's (idiosyncratic) language is a vast network of words and concepts. The meaning of any node in such a network—whether that node represents a sentence, a word, or a concept—is its location in the entire network, and thus depends on the meanings of all other sentences, words, and concepts in it.[14] (See §§3.5, 4.1.5.1, 7, below.) Beck et al. (1984) said, "through extensive reading, ... familiar words are encountered in new and varied contexts and each new context is a potential new facet of that word's network". In this way, a meaning for a word can depend on its context: (1) intended meanings of polysemous words can be determined by context (as in word sense disambiguation; see §4.2.2.2), and (2) a new context can enrich, extend, or change a meaning for a word.

3.2 The Nature of Context for CVA. By 'context', most CVA researchers in all disciplines have in mind *written* contexts as opposed to *spoken* contexts or visual or "situative" information (speaker, location, time, etc.). Still, there

---

[12]WJR's translation. Austin translates '*Satz*' as 'proposition', but, in this context, '*Satz*' must mean "sentence", because Frege is talking about words, not concepts. Frege wrote this before he made his *Sinn-Bedeutung* distinction, so it's at least unclear and probably unlikely that '*bedeuten*' meant "refer" here; cf. Austin's footnote on p. IIe of Frege 1884. See Resnik 1967, Pelletier 2001.

[13]" 'Sake' can be learned only contextually" (Quine 1960: 17; cf. pp. 14, 236).

is ambiguity (Engelbart & Theuerkauf 1999): Many researchers say that the reader can infer/guess/figure out, etc., the meaning of a word *from context* (Werner & Kaplan 1952, McKeown 1985, Schatz & Baldwin 1986). Sometimes they *say* that, but *mean*: "... from context *and* the reader's background knowledge" (Granger 1977, possibly Sternberg et al. 1983, Sternberg 1987, Hastings & Lytinen 1994ab). Sometimes, instead of talking about two, independent things—"context *and* background knowledge"—they talk about a unified thing: "context *including* background knowledge" (Nation & Coady 1988, Graesser & Bower 1990). But whereas 'context' as used in these studies connotes something in the *external world* (in particular, in the text containing the word), 'background knowledge' connotes something in the reader's *mind* (as in Clarke & Nation's step 1, §2.4.2, above). What *is* the context in *contextual* vocabulary acquisition?

Hirst (2000) is justifiably skeptical of attempts to pin down 'context'. But his skepticism is based on the widely different uses that the term has in different disciplines (such as knowledge representation and natural-language understanding), exacerbated by formal investigations that take the term as primitive (McCarthy 1993). He

[14]"[A] word's full concept is defined in the model memory to be all the nodes that can be reached by an exhaustive tracing process, originating at its initial, patriarchical type node, together with the total sum of relationships among these nodes specified by within-plane, token-to-token links" (Quillian 1967).

points out that anaphora is "interpreted with respect to the preceding text, ... so any preceding text is necessarily an element of the context." And then he observes that the sky's the limit: Context can "include just about anything in the circumstances of the utterance, and just about anything in the participants' knowledge or prior or current experience" (Hirst 2000, §4). Our point will be that, when it comes to CVA, the sky *must* be the limit (§4.1.5.1).

Our CVA software contains a clue to the nature of context as we need it: We represent, in a *single* semantic-network knowledge base, *both* the information in the text *and* the reader's background knowledge (Rapaport & Ehrlich 2000; cf. §3.4, below). This suggests that the relevant "context" for CVA is (at least a subpart of) the entire surrounding *network* of the unknown word. Before being more precise, we need to discuss "prior knowledge".

3.3 Prior Knowledge. The reader's "prior knowledge" is the "knowledge" that the reader has when *beginning* to read the text *and* that can be recalled as needed while reading. Because some of what readers think they know is mistaken, 'belief' or 'information' are more appropriate terms than 'knowledge'; so it must be understood that prior "knowledge" need not be true. "Prior" knowledge suggests the reader's knowledge *before* reading, i.e., beliefs brought to the text and available for use in understanding it. As we will see below, the reader's prior knowledge might have changed by the time the reader gets to *X* (and its immediately surrounding co-text), because it may "interact" with the text, giving rise to new beliefs. This is one of the principal components of reading comprehension. Similar terms used by other researchers have slightly different connotations:

1. '*Background* knowledge' lacks the temporal connotation but is otherwise synonymous. It might, however, more usefully refer to the information that the text's *author* assumes that the reader should have (Ong 1975). We could then distinguish the background knowledge *necessary* (or assumed) for understanding the text from the reader's actual prior knowledge. The author "is counting on ... words [in the text] evoking somewhat the same associations in the reader's mind as are present in his own mind. Either he has some sense of what his readers know or he assumes that their knowledge stores resemble his" (Simon 1996).
2. '*World* knowledge' connotes general knowledge about things *other* than the text's topic.
3. '*Domain* knowledge' is specialized, subject-specific knowledge about the text's topic.
4. '*Commonsense* knowledge' connotes the culturally-situated beliefs "everyone" has (e.g., that water is wet, that dogs are animals, that Columbus "discovered" America in 1492, etc.) but no specialized "domain" knowledge. We include under this rubric both the sort of very basic commonsense information that the CYC

knowledge-representation and reasoning system is concerned with (Lenat 1995)[15] and the somewhat more domain-specific information that the "cultural literacy" movement is concerned with (Hirsch 1987, 2003).

These notions overlap: The reader's prior knowledge includes much commonsense knowledge, and the author's intended background knowledge might include much domain knowledge. Reading comprehension can suffer when the reader's prior knowledge differs from the author's background knowledge.

Prior "knowledge" also includes expectations when encountering unknown words in specific reading situations: In an SAT exam, readers (should!) expect technical or obscure meanings. When reading a comic book, if 'kryptonite' is the unknown word, then readers should know that it is more likely to be a nonce word than an obscure geological one. When reading Shakespeare, readers should know that no unfamiliar word will refer to any modern contrivance such as email, but more likely to something from Elizabethan times. When reading a recipe, unknown words often refer to food ingredients ('cumin') or cooking methods ('braising').[16]

3.4 The Proper Definition of 'Context'. We use the expression **unfamiliar term** (denoted by '*X*') for a word or phrase that the reader has never seen before or has only the vaguest idea about its meaning, or that is being used in an unfamiliar way (cf. Kibby 1995). A **text** will be a (written) passage: a single sentence or an entire book (in a novel, knowledge of characters, settings, and themes might all be needed for CVA), usually containing *X*. *Pace* Schatz & Baldwin 1986, the text should not be *limited* to a 3-sentence window around *X* (§4.2.2.3, below).

A possibly awkward term that can serve a useful role is the **co-text of** *X* **as it occurs in** text *T*: the entire text *T* "minus" *X* (i.e., the entire text surrounding *X*).[17] So, if *X* = 'brachet', and if *T* is:

(T1) There came a white hart[18] running into the hall with a white **brachet** next to him, and thirty couples of black hounds came running after them. (Malory 1470: 66.)

then the co-text of 'brachet' as it occurs in T1 is:

There came a white hart running into the hall with a white _____ next to him, and thirty couples of black hounds came running after them.

---

[15]CYC is an "en*cyc*lopedic" knowledge-representation and reasoning system that attempts to encompass all commonsense information that is needed for general understanding [http: //www.cyc.com/].
[16]Goldfain, personal communication.
[17]The term seems to originate with Catford (1965: 31n2). Halliday (1978: 133) cites Catford, and Brown & Yule (1983: 46–50) cite Halliday; cf. Haastrup 1991, Widdowson 2004.
[18]We assume the reader knows that a hart is a deer, so that knowledge of harts can be used to help define 'brachet'. Readers who don't know that would have to try to figure out *its* meaning *before* that of 'brachet'.

The underscore marks the missing *X*'s location. Co-texts are used in "cloze" tests, in which a passage with a missing word is presented to a subject, who must then "fill in the blank", e.g., determine what that word might have been (Taylor 1953). In CVA, the reader is not usually trying to *find* a known-but-missing *word* (a "binary" task: one either succeeds or else fails). Rather, the reader is hypothesizing a *meaning* for a visible-but-unknown word (a "continuous" task: one can do well, poorly, or anywhere in between).[19]

The context of *X* for reader *R* is not merely *X*'s co-text, but rather a special kind of combination of *X*'s co-text with *R*'s prior knowledge. To take a simple example, after reading text T2:

(T2) Then the hart went running about the Round Table; as he went by the sideboard, the white brachet bit him in the buttock …. (Malory 1470: 66.)

most subjects infer that brachets are (probably) animals.[20] But they do not infer this solely from textual premise T2, because "every linguistic representation of some circumstance is in principle incomplete and *must be supplemented from our knowledge* about the circumstance" (Bühler 1934, our emphasis; cited by Kintsch 1980). I.e., *they must use an "enthymematic" premise from their prior knowledge* (Singer et al. 1990; cf. Anderson 1984, Suh & Trabasso 1993, Etzioni 2007), namely: If *x* bites *y*, then *x* is (probably) an animal. (Actually, it's more complex: We don't want to infer merely that this particular white brachet is an animal, but that brachets in general are animals.)

Two claims were just made: that an enthymematic premise is needed *and* that it comes from prior knowledge. An enthymematic premise is a "missing premise" that needs to be added to an argument to make it valid. Singer et al. 1990 call these "bridging inferences" connecting the text and the reader's prior knowledge. They do need to be inferred, though the inference involved is not (necessarily) deductive; rather, it is an "abductive" inference to the best explanation.[21] Thus, a reader might read in the text that a brachet bit a hart, abductively infer from prior knowledge

---

[19]Taylor invented cloze to help measure readability, not to do CVA. He preferred "[s]coring as correct only those fill-ins that precisely matched original words vs. *the more tedious process of judging synonyms and allowing half for each 'good enough' one*" (pp. 421f, our emphasis). His experiments suggested that "the more tedious method of judging synonyms as 'good enough' to be allocated half-counts yielded slightly larger total scores for the passages, but the degree of differentiation was virtually identical to scoring only precise matches" (p. 425). One conclusion is that being precise (which is easier to measure) suffices. Another is that *readability* measures are not altered by allowing "synonyms", but allowing them might better demonstrate *comprehension*. Cf. Kibby 1980.

[20]They also infer (unconsciously?) that 'brachet' is a noun, with plural 'brachets' (Goldfain, personal communication).

[21]The general form of abduction is the *deductive* fallacy of "affirming the consequent" (circumstantial evidence): From *P* implies *Q*, and *Q* is observed, infer that *P* might have been the case; i.e., *P* can explain the observation *Q*, so perhaps *P* is the case. Cf. Hobbs et al. 1993.

that if *x* bites *y*, then *x* is probably an animal, and then *deductively* infer from prior knowledge together with textual information that a brachet is probably an animal. The missing premise might come from prior knowledge or be found among, or deductively inferred from, information in the surrounding text. But in every situation that we have come across, at least one missing premise does, indeed, come from the reader's prior knowledge.

So, "context" is a combination of information *from the text* and information *in the reader's mind*. The "(co-)text" is in the external world; "prior knowledge" is in our minds. But, when you read, you "internalize" the text you are reading, i.e., you "bring it into" your mind (Gärdenfors 1997, 1999ab; Jackendoff 2002, 2006; Rapaport 2003a). Moreover, *this "internalized" text is more important than the external words on paper*. Here is a real-life example: One of us (WJR) read the sign on a truck parked outside our university cafeteria, where food-delivery trucks usually park, as "Mills Wedding and Specialty Cakes". Why had he never heard of this local bakery? Why might they be delivering a cake? Re-reading the truck's sign more carefully, it actually said, "Mills *Welding* and Specialty *Gases*"! What matters for understanding the text is not the *actual* text, but what you *think* it is.

So, let us resolve this "mind-body" duality by saying that the context of an unfamiliar term *X* for a reader *R* is a special combination of *R*'s *internalized* co-text of *X* with *R*'s prior knowledge. But what kind of combination? An active reader will typically make some (possibly unconscious) inferences while reading. E.g., from this small bit of text: "John went to the store. He bought a book.", readers will automatically infer that 'he' refers to John (some say that 'he' and 'John' both refer to the same person, others that the word 'he' refers back to the word 'John'—these differences don't matter for our purposes) and may automatically infer that John bought the book in the store that he went to. Or a reader of the phrase 'a white brachet' might infer (from prior, commonsense knowledge that only physical objects have color) that the brachet has a color or even that brachets are physical objects (Ehrlich 1995, Rapaport & Kibby 2007). Similarly, a reader might infer that, if a knight picks up a brachet and carries it away, then the brachet (whatever 'brachet' might mean) must be small enough to be picked up and carried (Ehrlich 1995, Rapaport & Kibby 2007). In these cases, the whole is *greater* than the sum of the parts: The combination of prior knowledge with internalized text might include some extra beliefs that are neither in the text nor previously in prior knowledge, i.e., that were not previously known—i.e., you can learn from reading! But the whole might also be *less* than the sum of the parts: From reading, you can also learn that one of your prior beliefs was *mistaken*. In that case, you revise your beliefs by *eliminating* something. Both ways of integrating text and prior knowledge are components of reading comprehension.

"Belief revision" is the subfield of AI, knowledge representation, and philosophy that studies this (Alchourrón et al. 1985, Martins & Shapiro 1988, Martins 1991, Gärdenfors 1992, Hansson 1999, Johnson 2006). The combination of the reader's prior knowledge and internalized (co-)text produces an updated, mental knowledge-base that is a "belief-revised integration" of the inputs. As the text is read, some passages from it will be "*added*" to the reader's prior knowledge, and perhaps new inferences will be drawn, "expanding" the prior knowledge base. Other text passages will be added, followed by the *elimination* of prior beliefs that are inconsistent with it (elimination is restricted to prior beliefs, because a reader typically assumes that the text is correct; Rapaport 1991, Rapaport & Shapiro 1995, 1999); this is called 'revision'. A few text passages (e.g., those involving typographical errors) might be added, then rejected when seen to be inconsistent with prior knowledge; this is called 'semi-revision'. Beliefs that are removed are said to be 'retracted'; such '*con*traction' of a knowledge base might also result in the *re*traction of other beliefs that inferentially depended upon the removed one (Rapaport 1991, Rapaport & Shapiro 1995). After finishing the text, the reader might consider all (relevant) beliefs in his or her newly expanded mental knowledge base, make new inferences, and eliminate further inconsistencies (such elimination is called 'consolidation'; Hanson 1999). Call the end result the '(belief-revised) integration' of the two inputs.

INSERT FIGURE 1 NEAR HERE

Figure 1: A belief-revised, integrated knowledge base and a text.

Pictorially, it might look like Figure 1. The left-hand rectangle represents the computational system's knowledge base or the reader's mind; initially, it consists of (say) four propositions representing the reader's prior knowledge: PK1, …, PK4. The right-hand rectangle represents the text being read; initially, it is empty (representing the time just before reading begins). At the next time step, the first sentence (T1) of the text is read. At the next time step, the reader "internalizes" T1, adding the (mental) proposition I(T1) to the "integrated" knowledge base. Here, "I" is an internalization function, encoding most of the processes involved in reading the sentence, so I(T1) is the reader's internalization of T1. At the next time step (or possibly as part of the internalization process), the reader might draw an inference from I(T1) and PK1, concluding some new proposition, P5, which becomes part of the "belief-revised" integrated knowledge base. Next, T2 is read and internalized as I(T2), with perhaps a new inference to P6, and similarly for T3 and I(T3). I(T3), however, might be inconsistent with prior belief PK4, and the reader might decide

to reject PK4 in favor of I(T3) (i.e., to temporarily—at least, while reading—or permanently stop believing PK4). Similarly, upon reading further sentences of the text, other prior beliefs (e.g., PK3) might be rejected and other inferences might be drawn (e.g., P7 from PK1 and PK2).

"Contextual" reasoning is done in the "context" *on the left-hand side*, i.e., the belief-revised, integrated knowledge base—i.e., *the reader's mind*. The context for CVA does not consist solely of the text being read (i.e., the co-text of the unfamiliar word) or of that (co-)text (merely) conjoined with the reader's prior knowledge. Rather, it is the reader's internalization of the (co-)text *integrated via belief revision* with the reader's prior knowledge.

One final detail: Because everything else has been internalized, we need a mental counterpart for the unfamiliar term in the text—an "internalized *X*". So, our final definition of 'context' for CVA makes it a three-place relation among a reader, a term, and a text:

Definition.

Let T be a text. Let R be a reader of T. Let X be a term in T that is unfamiliar to R. Let T—X be X's co-text in T.

Then: **the context that *R* should use to hypothesize a meaning for *R's* internalization of *X* as it occurs in *T***

$=_{\text{def}}$ *the belief-revised integration of R's prior knowledge with R's internalization of T—X.*

In plain English: Suppose that you have a text, a reader of that text, and a term in the text that is unfamiliar to the reader. Then the context that the reader should use in order to hypothesize (i.e., to figure out, or compute) a meaning for the reader's understanding of that word as it occurs in the text is the single, mental, knowledge-base resulting from the belief-revised integration of the reader's prior knowledge with the reader's internalized (co-)text.

3.5 Discussion. Our interpretation of current CVA instructional materials and methods is that they privilege text as the source of all clues to an unknown word's meaning. And too much, if not all, CVA instruction assumes that the author has (or has not) placed specific clues in the text to help readers determine a meaning for specific words in the text. This perspective overlooks more significant and useful CVA processes: text comprehension, prior knowledge brought to text comprehension by the reader, and the reasoning processes that combine them. Hobbs (1990) argues that a text's meaning is a function of both the text *and* the reader's mind.[22] Hence, a meaning for a word is not

[22]This may be a general cognitive principle: Whether an object is seen as white depends on the object in its environment *and* the perceiver's state of adaptation. If the perceiver has been in a red environment and visually adapted to red, then an object that looks white to her will look red to someone entering the red environment from a white environment (Webster et al. 2005).

usually given by the text alone. This view of the full context for CVA agrees with the experimental results of at least one reading researcher:

> Context has generally been assumed to refer to the immediate or local context that happens to surround a word. This conceptualization of context … does not take into account the mental representation that the reader is constructing on the basis of a variety of information contained in the text as well as prior knowledge. …
>
> The findings of this study point to the need to broaden our operationalization of context to include information that the reader has available in addition to information that is printed in close proximity to an unfamiliar word. In case the reader has been able to comprehend the text, then we must assume that *the amount of relevant information that the context provides is minimal when compared to the information contained in the mental representation*. (Diakidoy 1993: 3, 84–85; our emphasis.)

Our definition of 'context' meshes nicely with most cognitive-science and reading-theoretic views of text understanding as requiring schemata (e.g., scripts, frames, etc.; cf. Schank 1982, Rumelhart 1985), and also with most knowledge-representation and reasoning techniques in AI for processing text, which are, in turn, similar to Kintsch's (1988) construction-integration theory: The reader's mind is modeled by a knowledge base of "prior knowledge" expressed in a knowledge-representation language.

For us, that language is a semantic-network language (SNePS). As our computational cognitive agent reads the text, she (we have named "her" 'Cassie'; cf. Shapiro & Rapaport 1987, 1995; Shapiro 1989) incorporates the information in the text into her knowledge base, making inferences and performing belief revision along the way (using the SNePS Belief Revision system; Martins & Shapiro 1988, Martins & Cravo 1991, Johnson 2006). Finally, when asked to define one of the words she has read, she deductively searches this *single*, integrated knowledge base for information that can fill appropriate slots of a definition frame (for details, see Rapaport & Ehrlich 2000; "definition frames" are adapted from Van Daalen-Kapteijns & Elshout-Mohr 1981; the slots were inspired by Sternberg et al. 1983, Sternberg 1987).

As an example, consider the following series of passages (labeled "T") containing the unfamiliar word 'brachet' and the following prior knowledge (labeled "PK"):

T1 There came a white hart running into the hall with a white **brachet** next to him, and thirty couples of black hounds came running after them.

T2 As the hart went by the sideboard, the white **brachet** bit him in the buttock.

T3 The knight arose, took up the **brachet** and rode away with the **brachet**.

T4 A lady came in and cried aloud to King Arthur, "Sire, the **brachet** is mine."

T5 There was the white **brachet** which bayed at him fast. (Malory 1470: 66, 72; our boldface.)

PK1 Only physical objects have color.

PK2 Only animals bite.

PK3 Only small things can be picked up and carried.

PK4 Only valuable things are wanted.

PK5 Hounds are hunting dogs.

PK6 Only hounds bay.

Using these, Cassie outputs the following definition frame after processing ("reading") T1–T5:[23]

```
Definition of brachet:
Class Inclusions: hound, dog,
Possible Actions: bite buttock, bay, hunt,
Possible Properties: valuable, small, white,
```

This frame has three slots (`Class Inclusions`, `Possible Actions`, `Possible Properties`). The first has two fillers: a basic-level category (`dog`) and a category (`hound`) subordinate to "dog" and superordinate to "brachet". The second slot lists three actions that the only brachet the reader knows about is known to have performed (biting buttocks, baying, hunting); hence, these are considered to be "possible" actions of brachets in general (see §6.6, below). The third slot lists three similarly "possible" properties (`valuable`, `small`, `white`).

From our computational perspective, the "context" Cassie uses to hypothesize a meaning for a word consists of her prior knowledge together with that part of her knowledge base containing the information she integrated into it from the text. This matches our definition of 'context' for CVA. So, Cassie's definition is determined by relevant portions of the semantic-network knowledge-base. This is a version of a conceptual-role semantics that avoids Fodor & Lepore's (1992) alleged evils of holism (Rapaport 2002, 2003a).

Although Cassie reads in both a "bottom-up" (one sentence at a time) and a "top-down" fashion (using expectations based on her prior knowledge), she does not look back, scan ahead, or skip around, as human readers

[23]Updated definition frames are output after each passage containing the unknown word (Ehrlich 1995, Rapaport & Ehrlich 2000, Rapaport & Kibby 2007).

do.[24] But there is no reason in principle that she couldn't. Also, Cassie does not have to learn how to identify the printed form of words or match them to spoken forms, etc. These differences are not significant for our purposes, and our colleagues have investigated them computationally (e.g., Srihari et al. 2008).

4. How to Do Things with Words in Context[25]

Two often-cited papers by reading scientists (Beck et al. 1983; Schatz & Baldwin 1986) have claimed that some or most "natural", textual contexts are less than useful for doing CVA (as opposed to artificial, "pedagogical" textual contexts). However, their assumptions are inconsistent with our computational theory of CVA: It is possible to do *lots* of things with words in *any* (textual) context.

4.1 Are All Contexts Created Equal? In a paper subtitled "All Contexts Are Not Created Equal", Beck et al. (1983; cf. Beck et al. 2002) claim that "it is not true that every context is an appropriate or effective instructional means for vocabulary development" (177).[26] We argue, however, that *every* (textual) context contains *some* clues for constructing a meaning hypothesis. In this section, we examine where our approaches differ. (Except when quoting, 'textual context' will refer to the co-text surrounding an unfamiliar word, and 'wide context' to the reader's internalized co-text integrated with the reader's prior knowledge.)

4.1.1 The Role of Prior Knowledge. Beck et al. begin by pointing out that the co-text of a word "can give *clues* to *the word's meaning*" (177, our emphases). But a passage is not a clue for a reader without some other information—*supplied by the reader's prior knowledge*—that enables the reader to *recognize* it as a clue ('clue' is a relative term).[27] Nation (2001: 257, emphasis added) boasts that his guessing strategy "does *not* draw on background content knowledge" because "linguistic clues will be present in every context, background clues will not". But background (or prior) knowledge is essential and unavoidable, even in Nation's own strategy (§2.4.2, above): Where he says "Guess", he must mean "make an educated guess"—i.e., an inference that must rely on more premises than merely what is explicit in the text; such premises come from prior knowledge (see §3.4, above).

Prior knowledge introduces a great deal of variation into CVA in particular, and reading comprehension more generally, because readers bring to bear upon their interpretations of the text *idiosyncratic* prior knowledge (Dulin 1969, Garnham & Oakhill 1990, Rapaport 2003a):

---

[24]Karen Wieland, personal communication.
[25]With apologies to Austin 1962.
[26]Page references in §4.1 are to Beck et al. 1983, unless otherwise noted.
[27]Jean-Pierre Koenig, personal communication.

(1) The reader's internalization of the text involves interpretation (e.g., resolving pronoun anaphora) or immediate, unconscious inference (e.g., that 'he' refers to a male or that 'John' is a proper name typically referring to a male human) (cf. Garnham & Oakhill 1990: 383). Consider this natural passage:

> The archives of the medical department of Lourdes are filled with *dossiers* that detail well-authenticated cases of what are termed miraculous healings. (Murphy 2000: 45; our italics.)

Is this to be understood as saying (a) that the archives are filled with dossiers, and that *these dossiers* detail cases of miraculous healings? Or is it to be understood as saying (b) that the archives are filled with dossiers, and *dossiers in general* are things that detail cases of miraculous healings? The difference in interpretation has to do with whether "detail ... miraculous healings" is a restrictive relative clause (case (a)) or a non-restrictive relative clause (case (b)). Arguably, it should be understood as in (a); otherwise, the author should have written, 'The archives are filled with dossiers, *which* detail miraculous healings". But a reader (especially an ESL reader) might not be sensitive to this distinction (preferably indicated by 'that' *without* preceding comma vs. 'which' *with* preceding comma). The notion of misinterpretation cuts both ways: The author might not be sensitive to it, either, and might have written it one way though intending the other. It makes a difference for CVA. A reader who is unfamiliar with 'dossier' might conclude from the *restrictive* interpretation that a dossier is something found in an archive and that these particular dossiers detail miraculous healings, whereas a reader who internalized the *non*-restrictive interpretation might conclude that a dossier is something found in an archive that (necessarily) details miraculous healings. Our verbal protocols indicate that at least some readers of this passage do interpret it in the latter way.

(2) Even a common word can mean different things to different people: In some dialects of Indian English, upholstered furniture for sitting, even if it seats only one person, is a 'sofa', but an 'easy chair' or 'recliner' in American English. Thus, two fluent English speakers might interpret a passage containing the word 'sofa' differently: The text would be the same, but the readers' *internalized* texts would differ.[28]

(3) Variation also arises from *mis*interpretation (cf. Garnham & Oakhill 1990), even simple misreading (§3.4).

(4) Another source of variation stems from the amount of co-text that the reader can understand and therefore integrate into a mental model. Stanovich (1986: 370) notes that we must "distinguish the nominal context (what is on the page) from the effective context (what is being used by the reader)". Similarly, not all of the reader's prior knowledge may be consciously available at the time of reading.

---

[28]Shakthi Poornima, personal communication.

4.1.2 Do Words Have Unique Meanings? Beck et al.'s phrase 'the word's meaning' (177) or '*the* meaning of a word' suggests incorrectly (see §2.1) that:

Assumption A2 A word has a *unique* meaning.

To be charitable, we could say that what's normally intended by this definite description is "the meaning of a word *in the present context*" (recall Deighton's observation that context determines meaning; see §3.1, above). But from our observation that textual clues need to be supplemented with other information, it follows that the reader will supplement the co-text with idiosyncratic prior knowledge, and, consequently, each reader will interpret the word slightly differently. Of course, on this reading, Deighton is still essentially correct: Wide context determines *a* meaning for the word, but only further processing *reveals* that meaning.

4.1.3 Do Words Have Correct Meanings? A closely related limitation of current views of CVA is:

Assumption A3 A word has a *correct* meaning (in a given context),

and the associated notion that, if CVA does not result, at the time of application, in "the correct" meaning of the unfamiliar word, then CVA has failed. Beck et al. comment that "even the appearance of each target word in a strong, directive context [i.e., a context conducive to figuring out "a correct meaning"] is far from sufficient to develop *full knowledge* of word meaning" (180, our emphasis). This view defies not only commonsense about incremental learning, but also presumes that the sole purpose of CVA is vocabulary learning. In contrast, we argue that CVA is probably more useful to facilitate reading comprehension.

The most plausible interpretation of A3 is that there is a specific meaning that the *author* intended. However, we are concerned with a word's meaning as determined by the *reader's* internalized co-text integrated with the *reader's* prior knowledge. Our investigations suggest that it is almost always the case that the *author's* intended meaning is *not* thus determined. The best that can be hoped for is that a reader will be able to hypothesize or construct *a* meaning *for* the word (i.e., *give* or *assign* a meaning *to* the word), rather than figure out *the* meaning *of* (i.e., "belonging to") the word. "The meaning of things lies not in themselves but in our attitudes toward them" (St.-Exupéry 1948, cited in Sims 2003).

If the meaning that the reader figures out *is* the intended one, so much the better. If not, has the reader then *mis*understood the text? This is not necessarily bad: If no one ever misunderstood texts—or understood texts differently from others—then there would be little need for reading instruction, literary criticism, legal scholarship, etc. Because of individual differences in our idiosyncratic conceptual meanings, we *always misunderstand* each other

(Rapaport 2002, 2003a). Bertrand Russell celebrated this as the mechanism that makes conversation and the exchange of information possible:

> When one person uses a word, he does not mean by it the same thing as another person means by it. I have often heard it said that that is a misfortune. That is a mistake. It would be absolutely fatal if people meant the same things by their words. It would make all intercourse impossible, and language the most hopeless and useless thing imaginable, because the meaning you attach to your words must depend on the nature of the objects you are acquainted with, and because different people are acquainted with different objects, they would not be able to talk to each other unless they attached quite different meanings to their words. ... Take, for example, the word 'Piccadilly'. We, who are acquainted with Piccadilly, attach quite a different meaning to that word from any which could be attached to it by a person who had never been in London: and, supposing that you travel in foreign parts and expatiate on Piccadilly, you will convey to your hearers entirely different propositions from those in your mind. They will know Piccadilly as an important street in London; they may know a lot about it, but they will not know just the things one knows when one is walking along it. If you were to insist on language which was unambiguous, you would be unable to tell people at home what you had seen in foreign parts. It would be altogether incredibly inconvenient to have an unambiguous language, and therefore mercifully we have not got one. (Russell 1918: 195–196.)

The important question is not whether a reader can figure out *the correct* meaning of a word, but whether s/he can figure out *a* meaning for the word *sufficient to enable understanding the text well enough to continue reading*. Clarke & Nation (1980: 213)—more concerned with L2 understanding than with L1 vocabulary acquisition—note that "for a general understanding of a reading passage it is often sufficient to appreciate the general meaning of a word. ... Too often the search for a synonym ... meets with no success and has a discouraging effect." (Cf. Wieland 2008.) As Johnson-Laird (1987) has pointed out, we don't normally have, *nor do we need*, "full knowledge"—full, correct definitions—of the words that we understand: We can understand—well enough for most purposes—the sentence "During the Renaissance, Bernini cast a bronze of a mastiff eating truffles"[29] without being able to define any of its terms, as long as we have even a vague idea that, e.g., the Renaissance was some period in history, 'Bernini' is someone's name, "casting a bronze" has something to do with sculpture, bronze is some kind of (perhaps

[29]Johnson-Laird, personal communication; cf. Johnson-Laird 1983: 225, Widdowson 2004.

yellowish) metal, a mastiff is some kind of animal (maybe a dog), and truffles are something edible (maybe a kind of mushroom, maybe a kind of chocolate candy).

Consider the following passage from an article about contextual clues that can be taught in a classroom. (This might be a "pedagogical", not a "natural", passage, as defined in §4.1.4, below.)

> All chances for agreement were now gone, and compromise would now be impossible; in short, an *impasse* had been reached. (Dulin 1970; cf. Mudiyanur 2004.)

Here is one way a reader might figure out a meaning for 'impasse' from this text: From prior knowledge, we know that a compromise is an agreement and that if all chances for agreement are gone, then agreement is impossible. So both conjuncts of the first clause say more or less the same thing. Linguistic knowledge tells us that 'in short' is a clue that what follows summarizes what precedes it. So, to say that an impasse has been reached is to say that agreement is impossible. And that means that an impasse is a *dis*agreement.[30] Suppose that "deadlock" is "the correct meaning" (Waite 1998). If the reader decides that 'impasse' means "disagreement", not "deadlock", has the passage been misunderstood?

1. If the reader never sees the word 'impasse' again, it hardly matters whether the word has been "correctly" understood (though the reader has surely figured out *a* very plausible meaning).

2. If the reader sees the word again, in a context in which "disagreement" *is* a plausible meaning, then because the reader's prior knowledge now includes a belief that 'impasse' means "disagreement", this surely helps in understanding the new passage.

3. If the reader sees the word again, in a context in which "disagreement" is *not* a plausible meaning, but "deadlock" is—e.g., in a computer-science text discussing operating-system deadlocks, in which a particular deadlock is referred to as an "impasse"—then it *might* make little sense to consider the situation as a "disagreement", so:

   (a) The reader might decide that this occurrence of 'impasse' could not possibly mean "disagreement". Again, there are two possibilities:

[30]Plausibly, if agreement is *im*possible, then *dis*agreement *is* possible. If reaching a goal (albeit a negative goal, viz., an impasse—whatever that is) is *also* possible (perhaps because it has happened, and whatever happens is possible), then perhaps an impasse is also a *dis*agreement. These are defeasible inferences (subject to later rejection on the basis of new information), but our protocols show that readers make precisely these sorts of inferences.

i. She decides that she must have been wrong about 'impasse' meaning "disagreement" and now comes to believe (say) that it means "deadlock".

ii. She decides that 'impasse' is polysemous, and that "deadlock" is a second meaning. (Cf. Rapaport & Ehrlich 2000 on the polysemy of the verb 'to dress', which normally means "to put clothes on", but textual contexts such as "King Claudas dressed his spear before battle" suggest that to dress is *also* to prepare for battle.)

(b) Or the reader might try to reconcile the two possible meanings, perhaps by viewing deadlocks as disagreements, if only metaphorically (see CVA.7.2, in §5).

4.1.4 Two Kinds of Textual Context. Beck et al. are interested in using *textual* context to help *teach* "the" meaning of an unfamiliar word. We are interested in using *wide* context to help *figure out* "a" meaning for it, for the purpose of *understanding the text* containing it. These two interests don't always coincide (§2.4.3), especially if the former includes as one of its goals the reader's ability to *use* the word. That a given co-text might not clearly determine a word's "correct" meaning does not imply that a useful meaning cannot be figured out from it (especially because the wider context from which a meaning is figured out includes the reader's prior knowledge and is not therefore restricted to the co-text). Some co-texts certainly provide more clues than others. The question, however, is whether all CVA is to be spurned because of the less-helpful co-texts.

The top level of Beck et al.'s classification divides all (textual) contexts into *pedagogical* and *natural*. The former are "specifically designed for teaching designated unknown words" (178). Note that the only explicit example they give of a pedagogical co-text is for a *verb* (italicized below):

All the students made very good grades on the tests, so their teacher *commended* them for doing so well. (178)

By contrast, "the author of a natural context does not intend to convey *the meaning of a word* " (178, our emphasis). Note the assumptions about unique, correct meanings. In contrast, and following Deighton 1959 (see §3.1, above), we would say that the author of a natural co-text *does*—no doubt, unintentionally—convey *a* meaning for the word in question. Beck et al. go on to observe that natural "contexts will not necessarily provide *appropriate* cues to the meaning of a particular word" (178, our emphasis). This does not mean that no cues (or clues) are provided. It may well be that clues *are* provided for *a* meaning that helps the reader understand the passage. Finally, note that the pedagogical-natural distinction ultimately breaks down: A passage produced for pedagogical purposes

by one researcher might be taken as "natural" by another (see §4.1.6, below).

4.1.5 Four Kinds of Natural Co-Texts

4.1.5.1 Misdirective Co-Texts. Natural co-texts are divided into four categories. "At one end of our continuum are misdirective contexts, those that seem to direct the student to an *incorrect* meaning for a target word" (178, our emphasis). We agree that some co-texts are misdirective. But Beck et al.'s sole example is not clear cut:

> Sandra had won the dance contest and the audience's cheers brought her to the stage for an encore. "Every step she takes is so perfect and graceful," Ginny said *grudgingly*, as she watched Sandra dance. (178.)

Granted, a reader might incorrectly decide from this that 'grudgingly' meant something like "admiringly". But there are three problems with this example:

(1) No evidence is provided that this is, indeed, a natural co-text. This is minor; surely, many misdirective, natural co-texts could be found.

(2) If it is a natural co-text, it would be nice to see more of it. Indeed, another unwarranted assumption many CVA researchers make is this:

> **Assumption A4** (Textual) contexts have a fixed, usually small size.

But other clues—preceding or following the present, short co-text—might rule out "admiringly". Perhaps we know or could infer from other passages that Ginny is jealous of Sandra or inclined to ironic comments. From this passage, one could logically infer a disjunction of possible meanings of 'grudgingly' and later rule some of them out as more occurrences of the word are found (see §4.1.5.3, below).

(3) Most significantly, 'grudgingly' is an adverb. Now, another unwarranted assumption is this:

> **Assumption A5**. All words are equally easy (or equally difficult) to learn.

But adverbs and adjectives are notoriously hard cases not only for CVA but also for child-language (L1) acquisition (Granger 1977; Gentner 1981, 1982; Gillette et al. 1999; Dockrell et al. 2007: 579).

Thus, the evidence provided for the existence of misdirective co-texts is weak, because there should be *no* limit on the size of a co-text (see §4.2.2.3, below) and because the only example concerns an adverb, which can be difficult to interpret in any context. There *is* no "limit" on the size of the *wide* context. (This turns Hirst's (2000) criticism from a bug to a feature; see §3.2.) Certainly a reader's prior knowledge (which is part of that wide context) might include lots of beliefs that might assist in coming up with a plausible meaning for 'grudgingly' in this passage. Might a wider scope make it *harder* for the reader to identify relevant passages for CVA? We take a holistic view of

meaning: All passages are potentially relevant (Rapaport 2002). Our definition algorithms help filter out a dictionary-like definition from this wealth of data (Rapaport & Ehrlich 2000).

Another false assumption is also at work. Beck et al. conclude that "incorrect conclusions about word meaning are likely to be drawn" from misdirective co-texts (178). This assumes—incorrectly—that:

Assumption A6. Only one co-text can be used to figure out a meaning for a word.

Granted, if a word occurs only once, in the most grievous of misdirective co-texts, then it is likely that a reader would "draw an incorrect conclusion", if, indeed, the reader drew *any* conclusion. However, in such a case, it does not matter what, if anything, the reader concludes, for it is highly unlikely that anything crucial will turn on it. More likely, the reader will encounter the word again, and will have a chance to revise the initial meaning-hypothesis.

We agree that not all contexts are equally useful in a pedagogical situation for *learning* a meaning for a word (see §2.4.3, above). Natural texts—especially literary ones—are not designed for that purpose; yet they are likely the *only* contexts that readers will encounter in the real world. We are not seeking a foolproof method to learn meanings "indirectly"; the fastest and best way for a reader to learn an unknown word is to be told its meaning directly. But we are developing a method to assist readers in hypothesizing meanings in a way that facilitates independent reading.

In general, the task of CVA is one of hypothesis generation and testing; it is fundamentally a scientific task of developing a hypothesis (a theory about a word's meaning or possible meanings) to account for data (the text). It is not mere guessing (but cf. note 6). An alternative metaphor is that it is detective work: finding clues to determine, not "who done it", but "what does it mean" (Kibby et al., 2008; cf. Baumann et al. 2003: 462). And, like all hypotheses, theories, and conclusions drawn from circumstantial evidence (i.e., inferred abductively), it is susceptible to revision when more evidence is found.

All of this assumes that the reader is consciously aware of the unfamiliar word, notes its unfamiliarity, and remembers the word and its hypothesized meaning (if any) between encounters. Unfortunately, these assumptions are not necessarily the case. In real life, these are unavoidable problems. However, we expect that "word consciousness" grows with practice of CVA. In a classroom setting, these problems are less significant, because students can be made aware (or rewarded for awareness) of unfamiliar words, and subsequent encounters can be arranged to be close in time to previous ones.

4.1.5.2 Nondirective Co-Texts. The next category, "nondirective contexts, … seem to be of *no* assistance in directing the reader toward any particular meaning for a word" (178, our emphasis). Here is Beck et al.'s example:

> Dan heard the door open and wondered who had arrived. He couldn't make out the voices. Then he recognized the *lumbering* footsteps on the stairs and knew it was Aunt Grace. (178.)

Once again, there is no evidence of the sole example being natural, no mention of any larger co-text that might provide more clues, and the word is a modifier (this time, an adjective). Modifiers are hard to figure out from any context; it is not mis- or non-directive contexts that make them so (see §6.6, below).

The reader can ignore a single, unfamiliar word in *both* mis- *and* non-directive texts. But could an author use a word uniquely in such a way that it *is* crucial to understanding the text? Yes—authors can do pretty much anything they want. But, in such a case, the author would be assuming that the reader's prior knowledge includes the author's intended meaning for that word (recall Simon's quote in §3.3). As a literary conceit, it might be excusable; in expository writing, it would not be.

4.1.5.3 Syntactic Manipulation. Even a mis- or non-directive co-text can yield clues. A clue can be squeezed out of any co-text by syntactically manipulating the co-text to make the unfamiliar word its topic (its grammatical subject), much as one syntactically manipulates an equation in one unknown to turn it into an equation with the unknown on one side of the equals sign and its "co-text" on the other (cf. Higginbotham 1985, 1989; Rapaport 1986a). This technique can always be used to generate an initial hypothesis about a meaning for a word.

From the "misdirective" text in §4.1.5.1, we could infer that, whatever else 'grudgingly' might mean, it could be defined (if only vaguely) as "a way of saying something". Moreover, it could be defined (still vaguely) as "a way of (apparently) praising someone's performance". In both cases, we could list such ways, and hypothesize that 'grudgingly' is one of them. We put 'apparently' in parentheses, because readers who, depending on their prior knowledge, realize that sometimes praise can be given reluctantly or ironically might hypothesize that 'grudgingly' is that way of praising.[31] Similarly, from the "lumbering" passage, a reader might infer that lumbering is a property of footsteps, or footsteps on stairs, or even a *woman's* footsteps on stairs.

4.1.5.4 General Co-Texts. Not all co-texts containing modifiers are mis- or nondirective: "general contexts ... provide enough information for the reader to place the word in a general category" (178–179):

> Joe and Stan arrived at the party at 7 o'clock. By 9:30 the evening seemed to drag for Stan. But Joe really seemed to be having a good time at the party. "I wish I could be as *gregarious* as he is," thought Stan. (129.)

[31]Nation 2001: 235f makes similar points about Beck et al. in general and 'grudgingly', in particular.

Note that this adjective is contrasted with Stan's attitude. From a contrast, much can be inferred. Indeed, in our research, several adjectives that we have figured out meanings for occur in such contrastive co-texts: [32]

Unlike his brothers, who were noisy, outgoing, and very talkative, Fred was quite *taciturn*. (Dulin 1970.)[33]

From this, our CVA technique hypothesizes that 'taciturn' can mean "a personality characteristic of people who are not outgoing, talkative, or noisy, and possibly who talk little" (Lammert 2002). (Another example is our earlier discussion of 'tatterdemalion' in §2.2.)

4.1.5.5 Directive Co-Texts. Beck et al.'s fourth category is "directive contexts, which seem likely to lead the student to a specific, correct meaning for a word" (179). Their example is for a *noun*:

> When the cat pounced on the dog, he leapt up, yelping, and knocked over a shelf of books. The animals ran past Wendy, tripping her. She cried out and fell to the floor. As the noise and confusion mounted, Mother hollered upstairs, "What's all the *commotion*?" (179.)

Again, it's not clear whether this is a natural co-text.[34] More importantly, it may not be the co-text that is helpful as much as the fact that the word is a noun, which is generally easier to learn than adjectives and adverbs. Note, too, that this text is longer than the others, hence offers more opportunity for inferencing (see §4.2.2.3).

4.1.6 CVA, Neologisms, and Cloze. Beck et al. conducted an experiment involving subjects reading passages from basal readers. The researchers "categorized the contexts surrounding target words according to" their four-part "scheme", and they "then blacked out all parts of the target words, except morphemes that were common prefixes or suffixes.... Subjects were instructed to read each story and to try to fill in the blanks with the missing words or reasonable synonyms" (179). Several problems with this set-up may affect the results:

(1) The passages may indeed have been found in the "natural" co-text of a basal reader, but were the stories in these anthologies written especially for use in schools, or were they truly natural? (Remember: One reader's natural co-text might be another researcher's pedagogical one; see §4.1.4.)

---

[32]Another occurred in a (natural) co-text containing an equally useful, parallel construction: "In *The Pity of War* (1998), Ferguson argued that British involvement in World War I was unnecessary, far too costly in lives and money for any advantage gained, and a *Pyrrhic* victory that in many ways contributed to the end of the Empire" (Harsanyi 2003; cf. Anger 2003).

[33]This is probably not a natural co-text; it is what Beck et al. call "directive" (§4.1.3, above, and §4.1.5.5, below).

[34]The content, spelling ('leapt'), and the name 'Wendy' all suggest that this *might* be "natural" text from a version of *Peter Pan*. However, a very slightly different version, using the name 'Tonia', instead, appears in the National Institute for Literacy document "Put Reading First" [http://www.nifl.gov/partnershipforreading/publications/reading_first1vocab.html].

(2) How large were the surrounding co-texts? Recall that a small co-text might be nondirective or even misdirective, yet a slightly larger one might very well be directive.

(3) It is unclear whether the subjects were given any instruction on how to do CVA before the test. Here we find another unwarranted assumption:

Assumption A7. CVA "comes naturally", hence needs no guidelines, training, or practice.

But our project is focused on deliberate CVA, carefully taught and practiced.

(4) Another problem arises from the next unwarranted assumption:

Assumption A8. "Using context to guess the meaning of a semantically unfamiliar word is essentially the same as supplying the correct meaning in a cloze task" (Schatz & Baldwin 1986).

But this is not the case: In cloze tasks, a word is replaced with a blank, and the reader is invited to guess (rather than figure out) the unique, correct, missing word. But this is not CVA, whose goal is to figure out a meaning that is sufficient for understanding the passage. (Nor is cloze is valid as a measure of reading comprehension; Kibby 1980.)

Here, a serious methodological difficulty faces all CVA researchers: To find out if a subject can figure out a meaning for an unknown word from context, you don't want to use a word that the subject knows. You could filter out words (or subjects) by giving a pretest to determine whether the subjects know the test words. But then those who don't know them will have seen them at least once before (during the pretest), which risks contaminating the data. Finding obscure words (in natural co-texts, no less) that are highly unlikely to be known by any subjects is difficult; in any case, one might *want* to test familiar words. Two remaining alternatives—replace the word with a neologism or a blank—introduce complications: In our research on think-aloud protocols of students doing CVA (Kibby et al. 2004, Wieland 2008), we have found that, when students confront what they believe to be a real (but unknown) word, they focus their attention, thoughts, and efforts on meaning (i.e., what could this word mean?). However, neologisms, especially if particularly phony looking, will lead the subject to try to guess what the *original word* was rather than trying to figure out a *meaning* for it, e.g., a dictionary-like definition. A blank (as in a cloze test) even more clearly sends the message that the subject's job is to guess the *missing* (hence "correct") *word*. We are not alone in finding this a problem (Wolfe 2003, Gardner 2007), nor do we have any clever solutions. Our preferred technique for now is to use a plausible-sounding neologism (with appropriate affixes), to inform the subject that it is a word from another language that might or might not have a single-word counterpart in English, and to explain that the subject's job is to figure out what it might mean, not necessarily find an English synonym, exact or

inexact. (Translators sometimes leave untranslatable words in the original language, forcing the reader to do CVA; cf. Bartlett 2008: B9.)

4.1.7 Beck et al.'s Conclusions. Beck et al. claim that their experiment "clearly support[s] the categorization system" and "suggest[s] that it is precarious to believe that naturally occurring contexts are sufficient, or even generally helpful, in providing clues to promote initial acquisition of a word's meaning" (179). Significantly, "Only one subject could identify any word in the misdirective category" (179). This is significant, not because it supports their theory, but for almost the opposite reason: It suggests that CVA *can* be done even with misdirective co-texts, which supports *our* theory, not theirs.

They conclude that "Children most in need of vocabulary development—that is, less skilled readers who are unlikely to add to their vocabularies from outside sources—will receive little benefit from such indirect opportunities to gain information" (180–181). The false assumption underlying this conclusion is that:

Assumption A9. CVA can be of help only in vocabulary acquisition.

But another potential benefit far outweighs this: Because of high correlation between vocabulary knowledge, intellectual ability, and reading-comprehension ability, we believe that CVA strategies—if properly taught and practiced—can improve general reading comprehension. This is because the techniques that our computational theory employs and that, we believe, can be taught to readers, are almost exactly the techniques needed for improving reading comprehension: careful, slow reading; careful analysis of the text; a directed search for information useful for figuring out a meaning; application of relevant prior knowledge; and application of reasoning for the purpose of extracting information from the text. We are convinced that CVA has as least as much to contribute to reading comprehension in general as it does to vocabulary acquisition in particular. (For arguments and citations, see Wieland 2008, Ch. 1; see also Harris & Sipay 1990: 165.)

4.2 Are Context Clues Unreliable Predictors of Word Meanings? Schatz & Baldwin 1986 takes the case against context a giant step further, arguing "that context does not usually provide clues to the meanings of low-frequency words, and that context clues actually inhibit the correct prediction of word meanings just as often as they facilitate them" (440).[35]

4.2.1 Schatz & Baldwin's Argument. In summarizing the then-current state of the art, Schatz & Baldwin ironically note that "almost eight decades after the publication of ... [a] classic text [on teaching reading] ...,

[35]Page references in §4.2 to Schatz & Baldwin 1986, unless otherwise noted.

publishers, teachers, and the authors of reading methods textbooks have essentially the same perception of context as an *efficient* mechanism for inferring word meanings" (440, our emphasis). Given their rhetoric, the underlying, unwarranted assumption here appears to be:

Assumption A10. CVA is *not* efficient for inferring word meanings.

They seem to argue that *textual* context can't help you figure out "the" correct meaning of an unfamiliar word, so CVA is *not* "an effective strategy for inferring word meanings" (440). However, we argue that *wide* context *can* help you figure out *a* meaning for an unfamiliar word, so CVA *is* an effective strategy for inferring (better: figuring out) word meanings. As with Beck et al. 1983, note that the issue concerns the *purpose* of CVA. If its purpose is to get "the correct meaning", it is ineffective. But if its purpose is to get a meaning sufficient for understanding the passage in which the unfamiliar word occurs, it can be quite effective, even with an allegedly "misdirective" co-text.

Perhaps CVA is thought to be too magical, or perhaps too much is expected of it. Schatz & Baldwin claim that, "According to the current research literature, context clues should help readers to infer the meanings of ... [unfamiliar] words ... *without the need for readers to interrupt the reading act* with diversions to ... dictionaries, or other external sources of information" (441, our emphasis). This could only be the case if CVA were completely unconscious and immediate, so that one could read a passage with an unfamiliar word and instantaneously come to know what it means. This *may* hold for "incidental" CVA, but not for "deliberate" CVA. Our theory and our curriculum *require* interruption—not to access external sources—but for conscious, deliberate analysis of the co-text in light of prior knowledge. Computer models that appear to work instantaneously are actually doing quite a lot of active processing, which a human reader would need much more time for.

In any case, stopping to consult a dictionary does not suffice (see §1, above). CVA is the base case of a recursion, one of whose recursive clauses is "look it up in a dictionary". As Schwartz (1988: 111) points out, *CVA needs to be applied to the task of understanding a dictionary definition itself*, which is, after all, merely one more co-text containing the unfamiliar word (cf. Gardner 2007: 342).

4.2.2 Schatz & Baldwin's Methodology. Several experiments allegedly support their claims. But (their description of) their methodology is problematic:

4.2.2.1 Nouns and Verbs vs. Modifiers. Their first experiment took 25 "natural" passages from novels, selected according to an algorithm that randomly produced passages containing low-frequency words. They give only these examples (442–445, 448):

Adjectives	**Adverbs**	**Verb**	**Nouns**	**?**
glib	cogently	cozened	ameliorating	perambulating
imperious	ignominiously		coelum	
inexorable	ruefully		dearth	
pragmatic			yoke	
recondite				
salient				
waning				

Note that seven (or 47%) are adjectives, three (20%) are adverbs, one (7%) is a verb, four (27%) are nouns, and one ('perambulating') might be a noun, verb, or adjective, depending on the co-text. These are only "examples"; we are not given a full list of words, nor told whether these statistics are representative of the full sample. But, if they are, then fully two-thirds of the unfamiliar words are modifiers, known to be among the most difficult of words to learn meanings for. Of these, two of the nouns ('dearth', 'ameliorating') are presented as examples of words occurring in "facilitative" co-texts (448). Their example of a "confounding" co-text is for an adjective ('waning').

These examples raise more questions than they answer: What were the actual percentages of modifiers vs. nouns and verbs? Which lexical categories were hardest to determine meanings for? How do facilitative and confounding contexts correlate with lexical category? Schatz & Baldwin observe that, among "potential limitations" of their experiments, "a larger sample of words would certainly be desirable" but that their selection of "70 items ... offer[s] a larger and more representative sample than most studies of context clues" (449). But a representative sample of what? Of co-texts? Or of words? The sort of representativeness that is needed should (also) be a function of the variety of lexical category. What would happen with natural co-texts of, say, all four of Beck et al.'s categories, each containing nouns, verbs, adjectives, and adverbs (i.e., 16 possible types of co-text)? Schatz & Baldwin's and Beck et al.'s results may say more about the difficulty of learning meanings for modifiers (at least in short texts; see §4.2.2.3, below) than they do about weaknesses of contexts.

4.2.2.2 CVA vs. Word-Sense Disambiguation. Moreover, in two of their experiments, subjects were *not* involved in CVA. Rather, they were doing a related—but distinct—task known as "word-sense disambiguation" (WSD; Ide & Veronis 1998). The WSD task is to choose a meaning for a word from a given list of meanings; typically, the word is polysemous, and all items on the list are possible meanings for the word in different contexts.

The CVA task is to figure out a word's meaning "from scratch". WSD is a multiple-choice test, whereas CVA is an essay question (Ellen Prince, personal communication). In Schatz & Baldwin's experiment, the subjects merely had to replace the unfamiliar word with each multiple-choice meaning-candidate (each of which was a proposed one-word synonym) and see which of those five possible meanings fit better; CVA was not needed.

In the third experiment, real CVA was being tested. However, assumption A3 (about correct meanings) raises its head: "we were interested only in full denotative meanings or accurate synonyms" (446). There is no reason to expect that CVA will typically be able to deliver on such a challenge. But neither is there any reason to demand such high standards; once this constraint is relaxed, CVA can be seen to be a useful tool for vocabulary acquisition and general reading comprehension.

4.2.2.3 Space and Time Limits. Another assumption concerns the size of the co-text. The smaller the co-text, the less chance of figuring out a meaning, simply because there will be a minimum of textual clues. The larger the co-text, the greater the chance, simply because a large enough co-text might actually include a definition of the word! (Recall from §4.2.1 that CVA needs to be applied even in the case of an explicit definition!)

What is a reasonable size for a co-text? Our methodology has been to start small (typically, with the sentence containing the unknown word) and work "outwards" to preceding and succeeding passages, until enough co-text is provided to enable successful CVA. (Here, of course, 'successful' only means being able to figure out *a* meaning enabling the reader to understand enough of the passage to continue reading; it does *not* mean figuring out "the correct meaning of" the word.) This models what readers can do when faced with an unfamiliar word in normal reading: They are free to examine the rest of the text for possible clues. Schatz & Baldwin's limit on co-text size to 3 sentences (typically, the preceding sentence, the sentence containing the unfamiliar word, and the succeeding sentence) is arbitrary and too small. Any inability to do CVA from such a limited co-text shows at most that such co-texts are too small, not that CVA is unhelpful.

Another issue concerns time limits. Schatz & Baldwin do not tell us what limits were set, but do observe that "All students finished in the allotted time" (443). But real-life CVA has no time limits (other than self-imposed ones), and CVA might extend over a long period of time, as different texts are read.

4.2.2.4 Teaching CVA Techniques. Finally, there was no prior training in how to use CVA: "we did not control for the subjects' formal knowledge of how to use context clues" (449). Their finding "that students either could not or chose not to use context to infer the meanings of unknown words" (444) ignores the possibilities that the

students did not know that they *could* use context or that they did not know *how* to. Granted, "incidental" (or unconscious) CVA—the best explanation for how we learn most of our vocabulary—is something that we all do (see §2.3, above). But "deliberate" (or conscious) CVA is a skill that, while it may come naturally to some, can—and needs—to be taught, modeled, and practiced.

Thus, Schatz & Baldwin's conclusion that "context is an ineffective or little-used strategy for helping students infer the meanings of low-frequency words" (446) might only be true for untrained readers. It remains an open question whether proper training in CVA can make it an effective addition to the reader's arsenal of techniques for improving reading comprehension (for positive evidence, see Fukkink & De Glopper 1998, Kuhn & Stahl 1998, Swanborn & De Glopper 1999, Baumann et al. 2002).

Schatz & Baldwin disagree:

> It is possible that if the subjects had been given adequate training in using context clues, the context groups in these experiments might have performed better. We think such a result would be unlikely because the subjects were normal, fairly sophisticated senior high school students. If students don't have contextual skills by this point in time, they probably are not going to get them at all. (449.)

But how would they have gotten such skills if no one ever taught them? Assumption A7 (that CVA needs no training) is at work again. Students are not going to get "contextual skills" if they are not shown the possibility of getting them. Moreover, the widespread need for, and success of, critical thinking courses—not only at the primary- and secondary-school levels, but also in post-secondary education—strongly suggests that students need to, and can, be educated on these matters. How early can, or should, it be taught? In principle, the earlier, the better. But this is an open, empirical question.

4.2.3 Three Questions about CVA. Schatz & Baldwin ask three questions (447): (1) "Do traditional context clues occur with sufficient frequency to justify them as a major element of reading instruction?" This is irrelevant under our conception of CVA *if* CVA can be shown to foster good reading comprehension and critical-thinking skills. For clues need not occur frequently in order for the techniques for using them to be useful general skills. We believe that CVA can foster improved reading comprehension, but much more research is needed. Our answer is: Both "traditional" (§2.4) and other (§6.6) context clues are justified as a major element of reading instruction—as long as they are augmented by the reader's prior knowledge and by training in the application of reasoning abilities to improve text comprehension.

(2) "Does context *usually* provide accurate clues to the denotations and connotations of low-frequency words?" "Accuracy" is also irrelevant. Moreover, a "denotation" (i.e., an external referent of a word) is best provided by demonstration or by a graphic illustration, and a "connotation" (i.e., an association of the unfamiliar word with other (familiar) words) is not conducive to the sort of "accuracy" that Schatz & Baldwin (or Beck et al.) seem to have in mind. Our answer is: Context *can* provide clues to revisable hypotheses about an unfamiliar word's meaning.

(3) Are "difficult words in naturally occurring prose ... usually amenable to such analysis"? Yes; such words are always amenable to yielding at least some information about their meaning. At the very least, the information can be the "algebraic" meaning obtained by rephrasing the context to make the word its subject (§4.1.5.3). But the word will also have a meaning partly determined by the reader and the reader's accessible—and time-dependent—prior knowledge. Thus, every time you read a word in a text, it will have a meaning for you determined by the text integrated with your accessible prior knowledge at that time. Of course, none of the meanings that the word has for you is necessarily "the" meaning (in either a dictionary sense or that of a reading teacher).

5. A Positive Theory of Computational CVA

Progress is often made by questioning assumptions (Rapaport 1982). We have questioned the assumptions underlying Beck et al.'s and Schatz & Baldwin's arguments and experiments that challenge CVA. Their papers are best read as asserting that, *given those assumptions*, CVA is not as beneficial as some researchers claim. We now present our theory's contrasting beliefs. The details of our computational implementation and algorithm-based curriculum project are discussed in §6 and §7, below, respectively.

CVA.1 Every context can give some clue (if only minimal) to word meaning.

CVA.2 The textual context of a word (its co-text) contains clues to a word's meaning that must be supplemented by the reader's prior knowledge in order for the reader to figure out a meaning.

CVA.2.1. There are no such things as "good", "bad", "misdirective", "non-directive", "general", or "directive" co-texts; the value of a co-text depends in part on the reader's prior knowledge and ability to apply it to clues, and in part on the presence (or absence) of *potential* clues.[36]

CVA.3 CVA is neither a cloze-like task nor word-sense disambiguation.

CVA.4 Co-texts can be as small as a phrase or as large as an entire book; there are no arbitrary limits.

---

[36]Goldfain, personal communication, suggests that a (good) dictionary (e.g., Sinclair 1987, but probably not Mish 1983) is a good co-text *simpliciter* and that *Finnegan's Wake* is a bad co-text *simpliciter*.

CVA.5 Given enough contexts, CVA "asymptotically" approaches a "stable" meaning for a word.

CVA.6 In any context—even directive and pedagogical contexts—a word can have more than one meaning. This implies CVA.7:

CVA.7 A word does not have a (single) correct meaning, not even in a pedagogical context.

CVA.7.1. A word does not *need* a correct meaning (nor does any such correct meaning need to be understood) in order for a reader to be able to understand the word (in context).

CVA.7.2. Even a familiar and well-known word can *acquire* a new meaning in a new context, so meanings are continually being extended (e.g., when words are used metaphorically; cf. Lakoff 1987; Budiu & Anderson 2001; Rapaport 2000, 2006a).

CVA.8 Some lexical categories are harder to figure out meanings for than others (nouns are easiest, verbs a bit harder, modifiers the hardest).

CVA.9 CVA is an efficient method for inferring word meanings (in the absence of direct teaching).

CVA.10 CVA can improve general reading comprehension.

CVA.11 CVA can (and should) be taught.

6. A Computational Model of CVA

6.1 An Example Protocol. Before presenting our computational theory, consider one published protocol of a (presumably secondary-school-aged) reader ("Marian") figuring out what an unknown word might mean (Harmon 1999). The excerpt below begins with the text containing the unfamiliar word 'conglomerate', and follows with a transcript of Marian's reasoning:

> "How has the professor's brilliant career developed?" Chee asked.
>
> "Brilliantly. He's now chief legal counsel of Davidson-Bart, which I understand is what is called a multinational conglomerate. But mostly involved with the commercial credit end of export-import business. Makes money. Lives in Arlington." (Hillerman, *A Thief of Time*, p. 126)

MARIAN: The word is *conglomerate*. "He's now chief legal counsel of Davidson-Bart, which I understand is what is called a multinational *conglomerate*. But most involved with the commercial credit end of export-import business. Makes money. Lives in Arlington." I guess it's probably a company.

RESEARCHER: Why do you say that?

MARIAN: Well, Davidson-Bart that's like a company or some place like that. And so she's talking about Davidson-Bart, so it's a multinational *conglomerate*. A company that's two nations or more than one nation [pause] maybe.

RESEARCHER: What gives you that idea?

MARIAN: Because it's multinational or national. So "multi" means more or more than one. And then national or nation and so *conglomerate* I guess would be a company or something.

RESEARCHER: What kind of company?

MARIAN: Probably a business.

Marian has unconsciously performed a fairly complex inference (possibly two, which she has conflated). Cassie (our computer system), however, has to do it "consciously", so we programmers have to explicitly tell her how (and generally enough so that she can use this method in other situations), and then we educators can turn around and teach that method explicitly.

Here is Marian's inference:

1. 'Davidson-Bart' is a proper name. (This is prior knowledge, not necessarily of that name in particular, but of names in general; people can recognize the general form of names, e.g., because of capital letters.)
2. 'Davidson-Bart' is not the name of a person. (From integrating prior knowledge of names and of syntax with the co-text: A *person* would be unlikely to have a chief legal counsel; even were there such a person, the text would have said "chief legal counsel *for* Davidson-Bart", not "of" Davidson-Bart. Also, Davidson-Bart is referred to with 'which', not 'who'.)
3. Therefore, 'Davidson-Bart' is the name of a company; i.e., Davidson-Bart is a company. (Inference from prior knowledge (1) of the kind of entity that might need legal counsel and (2) that names are likely to be of people, companies, or geographical entities; it's not a person and probably not a geographical entity; so, it's probably a company.)
4. Davidson-Bart is a conglomerate. (This is an inference from the text, which says "Davidson-Bart is a multinational conglomerate". Prior linguistic knowledge tells us that if something is a multinational conglomerate (whatever that means), then it's probably a conglomerate. This is defeasible: A toy gun is *not* a gun; an alleged murderer is not necessarily a murderer.)
5. Therefore, possibly, a conglomerate is a company.

Now, line (5) is Marian's conclusion. It is justified by the following defeasible inference rule:

For any *x*, *y*, and *z*:	
IF *x* is a *y*	(e.g., if Davidson-Bart is a company)
AND IF *x* is a *z*	(e.g., if Davidson-Bart is a conglomerate)
AND IF *z* is unknown	(e.g., we don't know what 'conglomerate' means)
THEN possibly *z* is a *y*	(e.g., possibly a conglomerate is a company)

This very common kind of inference rule occurs in lots of other CVA contexts (see §6.3, below).

'Multinational' is an adjective that most likely applies to a company (this is prior knowledge), and this fact independently confirms our conclusion that Davidson-Bart is a company (rather than a person). But the whole discussion of 'multinational' might really be a secondary issue not directly related to, or needed for, doing CVA on 'conglomerate'. Or it might be that Marian's discussion of 'multinational' tells her in another way (besides the inference about names) that Davidson-Bart is a company, and so she has conflated two separate lines of reasoning.

Suppose there are other ways to infer from this context that a conglomerate is a company, and suppose that Marian didn't use any of them. It could still be that our computational system could use them and that we could *teach* Marian how to use them. We now turn to how that can be done.

6.2 Kinds of Prior Knowledge. Cassie is supplied with an initial stock of beliefs that models the prior knowledge that a reader brings to bear on a text. (She might also have beliefs about *how to do* certain things, though so far we have not explored this in our CVA project. She might also have "mental images"; e.g., she might be able to mentally visualize what she reads. She also has "subconscious" (or "tacit") *linguistic* knowledge (see §6.5, below). Our SNePS knowledge-representation language has three ways to express prior knowledge: (1) "basic propositions", (2) "proposition-based rules", and (3) "path-based rules".

(1) Examples of *basic propositions* are "Someone is named John", "Someone is tall", "Someone likes someone (else)", "Some particular kind of thing belongs to someone", etc. (see §6.5, below). Basic propositions are expressed in English by (a) simple subject-predicate sentences (usually without proper names—that someone has a certain name is itself a basic proposition) and by (b) simple relational sentences. Basic propositions are the sorts of sentences represented in first-order logic by, e.g., atomic sentences of the form *Px* or *Rxy*, i.e., sentences that assert that an entity *x* has a property *P* or that entities *x* and *y* stand in relation *R*. Basic propositions are probably most easily characterized negatively: They are not "rules" (as described next).

(2) *Proposition-based rules* are primarily conditional propositions of the form "if *P*, then *Q*" and usually involve universally quantified variables (e.g., "for all *x*, if *Px*, then *Qx*"; i.e., for any entity *x*, if *x* has property *P*, then *x* also has property *Q*). The SNePS Inference Package, which is the source of *inference* rules, allows Cassie to infer from a proposition-based rule of the form, e.g., if *P*, then *Q*, and a (typically, basic) proposition of the form *P*, that she should believe *Q* (i.e., the "*modus ponens*" rule of inference).

(3) *Path-based rules* generalize the inheritance feature of semantic networks, enabling Cassie to infer that Fido is an animal, if she believes that Fido is a brachet, that brachets are dogs, and that dogs are animals, or to believe that Fido has fur, if she believes that animals have fur and that Fido is an animal. The difference between proposition-based and path-based rules roughly corresponds to the difference between "consciously believed" and "subconsciously believed" rules (Shapiro 1991). This is all a vast oversimplification, but will suffice for now.

6.3 Special Rules. We try to limit the wide variety of prior knowledge (§3.3) to propositions necessary for Cassie to understand all words in the co-text of the unknown word *X*. In fact, we often use even less than this, limiting ourselves to that prior knowledge about the co-text that our analysis indicates is sufficient for Cassie to compute the meaning of *X*. Although this risks making Cassie too "brittle", it allows us to demonstrate a minimal set of prior knowledge that can support a plausible meaning. Besides basic propositions (usually necessary *or* sufficient conditions about the crucial terms in the co-text), we need rules of a very special and general sort. We saw one example in §6.1; here are a few more:

1. IF *x* is a subclass of *z*, and *x* is a subclass of *y*, and *z* has the property "unknown", and *y* is a subclass of *w*,
   THEN, presumably, *z* is a subclass of *y*, and *z* is a subclass of *w*.
2. IF action *A* is performed by agent *y* on object *z*, and action *B* is performed by *y* on *z*,
   THEN, presumably, *A* and *B* are similar.
3. IF *x* does *A*, and *A* has the property *P*, and *x* does *B*, and *B* is unknown,
   THEN, possibly, *B* also has the property *P*.

Such rules are fairly abstract and general, perhaps abductive or analogical in nature, and certainly defeasible. They are, we believe, essential to CVA. (For examples of these rules in action, see, e.g., Lammert 2002, Anger 2003, Goldfain 2003, Mudiyanar 2004, Schwartzmyer 2004.)

6.4 Source of Prior Knowledge. How does Cassie get this prior knowledge? In practice, we give it to her, though once she has it, it can be stored ("memorized") and re-used. (Each experiment in Ehrlich 1995 incorporated all prior

knowledge from previous experiments.) In general, Cassie would acquire her prior knowledge in any of the ways that one learns anything: reading, being told, previous reasoning, etc., including some of it being "innate". Cassie's prior knowledge is always unique, as is each human reader's—in part, a product of what she has read so far. Sometimes, we give Cassie prior knowledge that, although not strictly needed according to any of the (informal) criteria mentioned above, is such that human readers have indicated that they have (and use), as shown by our protocol case studies. Thus, we feel justified in giving Cassie some prior knowledge, including rules, that human readers seem to use, even if, on the face of them, this knowledge seems unmotivated.

6.5 Format of Prior Knowledge. Armed with her prior knowledge, Cassie begins to read the text. We can input the text to a computational parser, which outputs a semantic representation of the text. Currently, the grammar used by the parser is implemented in a generalized augmented-transition-network formalism (Woods 1970, Shapiro 1982). The output consists of a semantic network in the SNePS knowledge-representation language. (When our full system is implemented, Cassie will read all texts in the manner to be described below. Currently, we hand-code the output of this part of the process.)

For ease of grammar development, we constrain the possible input sentences to a small set, including those listed below, extended as necessary. However, each extension requires a corresponding extension to the definition algorithms, in order to include the new sentence type. The main idea is to analyze complex sentences into the "basic" propositions shown in Table 1, so that the meaning of the complex sentence is the (combined) meanings of the "basic" propositions into which it gets analyzed. In each entry in Table 1, if *x* is a proper name, then we represent the sentence by two propositions, one of which is that something has the proper name *x* (Rapaport 2006b, 2009).

TABLE 1 GOES APPROXIMATELY HERE

Each sentence is parsed into its constituent propositions. A proposition that is not already in the network is assimilated into it and "asserted" (i.e., Cassie comes to "believe" it). Cassie then does "forward" inference on it, modeling a reader who thinks about each sentence as it is read. If any proposition matches the antecedent of any prior-knowledge rule, that rule will fire (sometimes it has to be "tricked" into firing—an implementation-dependent "feature" that sometimes proves to be a bug; see Shapiro et al. 1982). This is the primary means by which Cassie infers the new information needed to hypothesize a definition.

6.6 Defining Words. At any point, we can ask Cassie to define any word *X*, whether or not it occurs in the current text (though typically, of course, it will). If *X* is not in Cassie's lexicon (because she has never read *X*, not even in the current text), she will respond with "I don't know". If *X is* in her lexicon, then—in the default case—Cassie will "algebraically/syntactically" manipulate the only sentence containing *X* so that *X* becomes its subject. Next, she will search through the belief-revised, integrated knowledge base (the "wide context") for information that can fill the slots of a definition frame. This models the task that readers might do by thinking hard about what they know about the unknown word from having read the text and thought carefully about applicable prior knowledge. The search is "deductive": The system looks for relevant information and also draws inferences whenever possible. Thus, it is an active search, simulating active reading and thinking. Each of these steps is repeated for subsequent occurrences of *X*, until a stable definition is reached. (Ehrlich 1995, Rapaport & Ehrlich 2000.)

The noun algorithm deductively searches the knowledge base for the following information about the unknown thing expressed by the word *X*:

- basic-level class memberships (e.g., "dog", rather than "animal"; Rosch 1978, Mervis & Rosch 1981); if Cassie fails to find or infer any, she seeks most-specific-level class memberships; failing that, she seeks names of individuals (e.g., she might decide that she doesn't know what kind of thing a brachet is, but she might know that 'Fido' is the name of one);
- properties of *X*s (size, color, etc.); if she can't find properties that she believes are exemplified by all *X*s, then she seeks properties of individual *X*s—these are considered to be "possible properties" of *X*s in the sense that our known *X* exemplifies it, so it is "possible" that some *X*s exemplify them;
- structural information about *X*s (part-whole, physical structure, etc.); if she can't find structural information that she believes is exemplified by all *X*s, then she seeks "possible" structural information exemplified by individual *X*s;
- acts (or "possible" acts) that *X*s perform or that can be done to, or with, *X*s;
- agents that do things to, or with, *X*s, or to whom things can be done *X*s, or that own *X*s;
- possible synonyms and antonyms of the word *X*.

The verb algorithm deductively searches the knowledge base for:

- class-membership information (e.g., based on Schank & Rieger's (1974) "Conceptual Dependency" classification or Levin's (1993) cognitive-linguistic classification): What kind of act is *X*-ing? (e.g., walking is a kind of moving). What kinds of acts are *X*-ings? (e.g., sauntering is a kind of walking);
- properties or manners of *X*-ing (e.g., moving by foot, walking slowly);
- transitivity or subcategorization information (i.e., agents, direct objects, indirect objects, instruments, etc.);
- class membership information about the agents, direct objects, indirect objects, instruments, etc.;
- possible synonyms and antonyms of the word *X*;
- causes and effects of *X*-ing.

Adjectives and adverbs are very much more difficult to figure out meanings for unless there is very specific kinds of information in the context: After all, if 'car' is modified by an unknown adjective *X*, *X* could refer to the car's color, style, speed, etc. On the other hand, it is unlikely to refer to the car's taste (e.g., *X* is unlikely to be 'salty', though it could be 'sweet', used metaphorically). Thus, modifiers can be categorized in much the same way that nouns and verbs can; this information—together with such information as contrasting modifiers that might be in the context—can help in computing a meaning for *X*. In short, the system constructs a definition of word *X* in terms of some (but not all) other nodes that are directly or indirectly linked to the node representing *X*.

7. The Curriculum.

The original version of our computational system was based on an analysis of how the meaning of a node in a holistic semantic-network would depend on the other nodes in the network (Quillian 1967, Rapaport 1981). Later modifications have been based on protocols of human readers doing CVA (Wieland 2008; cf. Kibby 2007).

Cassie (our computer system) inevitably works more efficiently and completely than a human reader: She never loses concentration during reading. She has perfect memory, never forgetting what she has read or been told, and easily retrieving information from memory. And she is a (near-) perfect reasoner, inferring everything inferable from this information (at least in principle—there are certain implementation-dependent limitations).

Humans, however, get bored, are forgetful, and don't always draw every relevant logical consequence. After reading that a knight picked up a "brachet" and rode away with it, about half of the readers state that this gives them no useful information about brachets—until they are asked how big a brachet is. Then the proverbial mental lightbulb lights, and they realize that a brachet must be small enough to be picked up. They all knew that, if someone can pick something up, then the item must be relatively small and lightweight, but half of them either

forgot that or weren't thinking about it, failing to draw the inference until it was pointed out to them. Cassie always infers such things. The important point is not that the computer is "better" than a human reader (assuming that it has as much prior knowledge as the human reader), but that the computer simulates what a human reader *can* do.

Our simulation is implemented in a symbolic AI system, not a statistically-based or connectionist system (§1). Therefore, we can turn things around and have a human reader simulate Cassie! This emphatically does *not* mean that such a reader would be "thinking like a computer": rigidly, uncreatively, "mechanically". Rather, it means that *what we have learned by teaching our computer to do CVA can now be taught to readers who need guidance in doing it*. Clearly, there are things that the computer can do automatically and quickly but that a human might have to be taught how to do, or coaxed into doing. For instance, a computer can quickly find all class membership information about *X*, based simply on the knowledge-representation scheme; a human has to search his or her memory without such assistance. And there are things that a human will be able to do that our computer cannot (yet), such as suddenly have an "Aha!" experience that suggests a hypothesis about a meaning for *X*.

But we *can* devise a curriculum for teaching CVA that is a human adaptation of our rule-based algorithms. A statistically-based algorithm could not be so adapted: The students would first have to be taught elementary statistics and then shown a statistically-relevant sample of texts containing *X*. Our rule-based algorithms only require a single occurrence of *X* in a single text. However, some things need to be added to the curriculum to accommodate human strengths and weaknesses. And, although a computer must slavishly follow its own algorithm, a human reader must be allowed some freedom concerning which rules to follow at which times. Nevertheless, we can supply an algorithm that a human reader can always rely on when at a loss for what to do (no guessing or "miracles" needed). Finally, the human-oriented CVA strategies must be embedded in a "scaffolded" curriculum that (a) begins with examples and instructions provided by a teacher familiar with the technique, (b) followed by teacher-modeling with student participation—perhaps the students challenge the teacher with an unknown word in an unfamiliar text, or the students and teacher work as a team. (c) Next, the burden is placed on the students, who now take the lead with the teacher's help, (d) followed by small groups of students working together. (e) Finally, each student is given an opportunity to work on his or her own, so that the technique evolves into a tool that readers can rely on in future, independent reading. We call the full curriculum "Contextual Semantic Investigation", with the currently popular abbreviation "CSI" (Kibby et al., 2008), emphasizing one of our two guiding metaphors: detective work—the reader

must seek clues in the text, supplemented by his or her prior knowledge, to identify a hypothesis (a "suspect", to continue the detective metaphor) and then make a case for the suspect's "guilt" (i.e., the word's meaning).

7.1 The Basic, Human-Centered (Curricular) Algorithm: Generate and Test a Meaning Hypothesis

To figure out a meaning for an unknown word:

1. Become aware of the unknown word *X* and of the need to understand it.
2. Generate and test a meaning hypothesis, as follows:

Repeat: (a) Choose a textual context *C* to focus on

(b) *Generate a hypothesis H* about *X*'s meaning in the "wider" context consisting of *C* integrated with the reader's prior knowledge (§7.3)

(c) *Test H* (§7.2)

until *H* is a plausible meaning for *X* in the current "wide" context.

Step 7.1.1 is our first concession to human frailty: Because readers often skim right over an unfamiliar word (§1), the first step is to make the reader aware of it and to see the need for understanding it. This is harder to do when reading on one's own than in a classroom (where the teacher can simply tell the students that they need to figure out what the word means). But by the end of classroom instruction, readers should have become more aware of unfamiliar words. (For empirical support that this does occur, see Beck et al. 1982, Christ 2007.) Step 7.1.2(a) allows the reader to expand the textual context under consideration, if needed. The process ends when readers have hypothesized a meaning consistent with both the text and their prior knowledge.

7.2 Test the Hypothesis. If the hypothesis is not tested, then a poor understanding of a word's meaning can lead to further misunderstanding later in the text. Testing can be done by simple substitution:

To test H:

1. Replace all occurrences of *X* (in the sentence in which it appears) with *H*.
2. **If** the sentence with *X* replaced by *H* makes sense, **then** continue reading

else generate and test a new meaning-hypothesis (§7.1.2).

If, say, *X* is a noun, but the reader has construed its definition in verb form, then the substitution shouldn't make sense,[37] and the student will have to revise *H*, which should be fairly straightforward with the teacher's help.

[37]Our campus restaurant, The Tiffin Room, used to have, on the cover of its menu, a *faux* dictionary definition of its name that said something like: "**tiffin** (noun): to eat".

7.3 Generate a Hypothesis. The bulk of the work lies in *generating* a meaning hypothesis. Our curriculum differs from our computer algorithm by giving the reader a chance to make a guess. Computers can't (easily) guess (§2.4.2). We suspect that less-able readers can't, either. It is with them in mind that the rest of our algorithm goes into a great deal of detail. But some better readers might be able to guess, either intuitively or on the basis of prior knowledge of prefixes and suffixes; this is where we give them that opportunity. (There is no requirement, unlike in a typical serial computer, that these steps be done in any particular order, nor even that they all be done.)

To generate H:

1. Guess an "intuitive" *H* and test it.
2. **If** you can't guess an intuitive *H*, **or if** your intuitively-guessed *H* fails the test,

   then do one or more of the following, in any order:

   (a) **if** you have read *X* before **& if** you (vaguely) recall its meaning, **then** test that earlier meaning

   (b) **if** you can generate a meaning from *X*'s morphology, **then** test that meaning

   (c) **if** you can *make an "educated guess"* (§7.4), **then** test it

7.4 Make an Educated Guess. "Educated" (as opposed to "mere") guessing results from careful, active thinking:

To make an "educated" guess:

1. Summarize the entire text so far
2. Activate your prior knowledge about the topic
3. Re-read the sentence containing *X* slowly and actively
4. Determine *X*'s part of speech
5. Draw whatever inferences you can from: the text integrated with your prior knowledge
6. Generate *H* based on all this.

Step 7.4.2 might be accomplished by class discussion or instruction (Greene 2010: 28–29 has a useful suggestion along these lines). Steps 7.4.5 and 7.4.6 are intentionally vague. They are not part of our computer program (which forms the basis of §7.5, below) but are included in the curriculum as a guide for good readers who do not need the more detailed assistance in the next several steps.

7.5 The CVA Algorithm. What can the reader who has not succeeded in generating a hypothesis do?

1. **If** all previous steps fail, **then** do CVA:
   (a) *"Solve for X"* (§7.6)
   (b) *Search context for clues* (§7.7)
   (c) *Construct H* (§7.8)

7.6 "Algebraic" Manipulation. The first of these steps is "algebraic" manipulation (§4.1.5.3):

To "solve for X":

1. Syntactically manipulate the sentence containing *X* so that *X* is its subject
2. Generate a list of possible synonyms to serve as "hypotheses in waiting"

E.g., syntactically manipulate the sentence "A hart ran into King Arthur's hall with a *brachet* (*X*) next to him", like an algebraic equation, to yield: A brachet (*X*) is something that was next to a hart that ran into King Arthur's hall. A list of things that could be next to the hart might include: an item of furniture, a female deer, a dog, an animal, etc. Each of these could be tested as a possible meaning or could be held in abeyance until further evidence favoring or conflicting with one or the other was found in a later sentence.

7.7 Search for Clues. Next, the wide context (i.e., the reader's prior knowledge integrated with the reader's memory of what was read in the text) must be searched for clues.

To search the wide context for clues:

1. **If** *X* is a noun, **then** search the wide context for clues about *X*'s ...
   • class membership, properties, structure, acts, agents, comparisons, contrasts.
2. **If** *X* is a verb, **then** search the wide context for clues about *X*'s ...
   • class membership, what kind of act *X*ing is, what kinds of acts are *X*ings, properties of *X*ing (e.g, manner), transitivity, agents and objects of *X*ing, comparisons, contrasts.
3. **If** *X* is an adjective or adverb, **then** search the wide context for clues about *X*'s ...
   • class membership (is it a color adjective, a size adjective, a shape adjective, etc.?), contrasts (is it an opposite or complement of something else mentioned?), parallels (is it one of several otherwise similar modifiers in the sentence?)

7.8. Construct a Definition. Armed with all of this information, the reader now has to construct a meaning hypothesis. We suggest that the classical Aristotelian technique of definition by genus and differentia be combined with the definition-"map" strategy of Schwartz & Raphael 1985 (cf. Schwartz 1988).

To create H:

1. Express ("important" parts of) the definition frame in a single sentence by answering these questions, based on the results of the search in step 7.7:
   (a) What (kind of thing) is *X*?
   (b) What is it like? (What properties does it have? What relations does it stand in?)
   (c) How does it differ from other things of that kind?
   (d) What are some examples?

The sorts of sentences that we have in mind are the definition sentences used in the *Collins COBUILD* dictionary for speakers of English as a second language (Sinclair 1987). For example, the definition frame for 'brachet' (§3.5) can be expressed by the single-sentence definition: "A brachet is a hound (a kind of dog) that can bite, bay, and hunt, and that may be valuable, small, and white." (Whether being a dog is "important" to the definition probably depends on how familiar the reader is with the concept of a hound.)

8. Summary and Conclusion

CVA is a hard problem. It is part of the general problem of natural-language understanding, the computational solution for which is "AI-complete" (Shapiro 1992): solving it involves developing a complete theory of human cognition (solving all other AI problems). We have formalized a partial solution to the CVA problem in a computer program with two important features: It can be adapted as a useful procedure for *human* readers, and it can be taught in a classroom setting as a means for vocabulary acquisition and to improve reading comprehension. The curriculum is flexible and adaptable, not scripted or lock-step. But the steps are detailed and available for those who need them.

Many open research questions remain: Can the curriculum be taught successfully? At what levels? Does it need further modification to make it humanly usable? Does it help improve vocabulary? Reading comprehension? Critical-thinking skills? We are exploring these issues and hope that others will join us.

9. A Closing Anecdote

I close with a final Castañeda anecdote. After he passed away, I dreamed that he was spending his afterlife in a room all four of whose walls were filled with books, sitting at a table piled with papers, and happily typing away at his

computer, continuing to write philosophy. I was happy and consoled to know that he was doing what he loved most. I was disappointed that I would never get a chance to read those posthumous papers…but then Castañeda 1999 was published!

Acknowledgments

This research was supported in part by the National Science Foundation Research on Learning and Education Program, grant #REC-0106338, 7/1/2001–12/31/2003. Most of §3 (inter alia) is based on Rapaport 2003b, most of §4 (inter alia) on Rapaport 2005. We are grateful to Tanya Christ, Debra Dechert, Albert Goldfain, Michael W. Kandefer, Jean-Pierre Koenig, Shakthi Poornima, Michael J. Prentice, Stuart C. Shapiro, Karen M. Wieland, and the SNePS Research Group for comments and advice.

References

Aist, G. (2000). Helping children learn vocabulary during computer assisted oral reading. In *Proc. 17th Nat'l. Conf. Artif. Intell.* (pp. 1100-1101). Menlo Park, CA, & Cambridge, MA: AAAI/MIT Press.

Alchourrón, C.E.; Gärdenfors, P.; & Makinson, D. (1985). On the logic of theory change. *J. Symbolic Logic 50*, 510–530.

Ames, W.S. (1966). The development of a classification scheme of contextual aids. *Read. Res. Quart. 2,* 57–82.

Anderson, R.C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R.C. Anderson et al. (eds.), *Learning to read in American schools* (pp. 243-257). Erlbaum.

Anger, B. (2003). Defining 'Pyrrhic victory' from context" [http://www.cse.buffalo.edu/~rapaport/CVA/Pyrrhic/Pyrrhic-Victory-Report.pdf].

Associated Press (2008). "Deer with unicorn-like horn lives in nature preserve in Italy. *Buffalo News* (12 June).

Austin, J.L. (1962). *How to do things with words*. Oxford: Clarendon.

Bartlett, T. (2008). The betrayal of Judas. *Chronicle of Higher Education* (30 May), B6–B10.

Baumann, J.F.; Edwards, E.C.; Boland, E.M.; Olejnik, S., & Kame'enui, E.J. (2003). Vocabulary tricks. *American Educational Research Journal 40*, 447–494.

Baumann, J.F.; Edwards, E.C.; Front, G.; Tereshinski, C.A.; Kame'enui, E.J.; & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly 37*, 150–176.

Beals, D.E. (1997). Sources of support for learning words in conversation. *Journal of Child Language 24*, 673–694.

Beck, I.L.; McKeown, M.G.; & Kucan, L. (2002). *Bringing words to life*. New York: Guilford.

Beck, I.L.; McKeown, M.G.; & McCaslin, E.S. (1983). Vocabulary development: All contexts are not created equal. *Elementary School Journal 83,* 177–181.

Beck, I.L.; McKeown, M.G.; & Omanson, R.C. (1984). The fertility of some types of vocabulary instruction. Paper presented at American Educational Research Association (New Orleans, April).

Beck, I.L.; Perfetti, C.A.; & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology 74,* 506–521.

Blachowicz, C., & Fisher, P. (1996). Learning vocabulary from context. (pp. 15-34). *Teaching vocabulary in all classrooms*. Englewood Cliffs, NJ: Prentice-Hall/Merrill.

Bowey, J.A.; & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading. *Journal of Experimental Child Psychology 92:* 203–219.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Bruner, J. (1978). Learning how to do things with words. In J. Bruner & A. Garton (Eds.). *Human growth and development* (pp. 62-84). Oxford: Oxford U. Press).

Brysbaert, M.; Drieghe, D.; & Vitu, F. (2005). Word skipping. In G. Underwood (Ed.). *Cognitive processes in eye guidance* (pp. 53-78). Oxford Univ. Press.

Budiu, R., & Anderson, J.R. (2001). Word learning in context: Metaphors and neologisms. *Technical Report CMU-CS-01-147* (Carnegie Mellon University, School of Computer Science).

Bühler, Karl (1934). *Sprachtheorie*. Jena: Rischer Verlag.

Buikema, J.L., & Graves, M.F. (1993). Teaching students to use context cues to infer word meanings. *Journal of Reading 36*(6), 450–457.

Carroll, L. (1896). *Through the looking-glass and what Alice found there*. [http://www.cs.indiana.edu/metastuff/looking/lookingdir.html].

Castañeda, Hector-Neri (1972), Thinking and the structure of the world, *Philosophia 4* (1974), 3–40.

Castañeda, Hector-Neri (1999). *The Phenomeno-Logic of the I.* Ed. By J.G. Hart & T. Kapitan. Bloomington, IN: Indiana University Press.

Catford, J.C. (1965). *A linguistic theory of translation*. Oxford University Press.

Christ, T. (2007). Oral language exposure and incidental vocabulary acquisition. PhD dissertation (Buffalo: SUNY Buffalo Dept. of Learning & Instruction).

Church, A. (1956). *Introduction to Mathematical Logic*. Princeton, NJ: Princeton University Press.

Clancey, W.J. (2008). Scientific antecedents of situated cognition. In P. Robbins & M. Aydede (Eds.). *Cambridge handbook of situated cognition*. New York: Cambridge.

Clarke, D.F., & Nation, I.S.P. (1980). Guessing the meanings of words from context. *System 8:* 211–220.

Cole, David. (1999). Note on analyticity and the definability of "bachelor". Accessed 15 November 2011 from [http://www.d.umn.edu/~dcole/bachelor.htm].

Cunningham, A.E., & Stanovich, K.E. (1998). What reading does for the mind. *American Educator* (Spring/Summer): 1–8.

Deighton, L.C. (1959). *Vocabulary development in the classroom*. New York: Teachers College, 1970.

Denning, P.J. (2007). Computing is a natural science. *Communications of the ACM 50*(7) (July): 13–18.

Diakidoy, I.-A.N. (1993). The role of reading comprehension and local context characteristics in word meaning acquisition during reading. PhD dissertation. Urbana-Champaign: University of Illinois.

Dulin, K.L. (1969). New research on context clues. *J. Reading 13*: 33–38, 53.

Dulin, K.L. (1970). Using context clues in word recognition and comprehension. *Read. Teach. 23*: 440–445, 469.

Ehrlich, K. (1995). Automatic vocabulary expansion through narrative context. *Technical Report 95-09* (Buffalo: SUNY Buffalo Dept. Computer Science).

Ehrlich, K. (2004). Default reasoning using monotonic logic. In *Proceedings of the 15th Midwest AI & Cognitive Science Conference* (pp. 48–54).

Ehrlich, K., & Rapaport, W.J. (1997). A computational theory of vocabulary expansion. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 205-210). Mahwah, NJ: Erlbaum.

Ehrlich, K., & Rapaport, W.J. (2005). A cycle of learning. In *Proc. of the 26th Annual Conference of the Cognitive Science Soc.* (p. 1555). Erlbaum.

Elman, J.L. (2007). On words and dinosaur bones. In D.S. McNamara & J.G. Trafton (Eds.). In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (p. 4). Austin, TX: Cognitive Science Society.

Elman, J.L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science 33*(4), 547-582.

Engelbart, S.M., & Theuerkauf, B. (1999). Defining context within vocabulary acquisition. *Language Teaching Research 3*, 57–69.

Etzioni, O. (Ed.) (2007). *Machine reading*. Technical Report SS-07-06. Menlo Park, CA: AAAI.

Fodor, J.A. (1975). *The language of thought*. New York: Crowell.

Fodor, J.A.; & Lepore, E. (1992). *Holism*. Cambridge, MA: Blackwell.

Franklin, H.B. (2001, Sept.). The most important fish in the sea. *Discover 22*, 44–50.

Frege, G. (1884). *The foundations of arithmetic*, J.L. Austin (Trans.). Oxford: Blackwell, 1968.

Frege, G. (1892). On sense and reference, M. Black (Trans.). In P. Geach & M. Black (Eds.). *Translations from the philosophical writings of Gottlob Frege* (pp. 56-78). Oxford: Blackwell, 1970.

Fukkink, R.G., & De Glopper, K. (1998). Effects of instruction in deriving word meaning from context. *Review of Educational Research 68*, 450–458.

Gärdenfors, P. (Ed.) (1992). *Belief revision*. Cambridge University Press.

Gärdenfors, P. (1997). Meanings as conceptual structures. In M. Carrier & P. Machamer (Eds.). *Mindscapes* (University of Pittsburgh Press).

Gärdenfors, P. (1999a). Does semantics need reality? In A. Riegler, et al. (Eds.). *Understanding representation in the cognitive sciences*. Dordrecht: Kluwer.

Gärdenfors, P. (1999b). Some tenets of cognitive semantics. In J.S. Allwood & P. Gärdenfors (Eds.). *Cognitive semantics*. Amsterdam: Benjamins.

Gardner, D. (2007). Children's immediate understanding of vocabulary. *Reading Psychology 28,* 331–373.

Garnham, A., & Oakhill, J. (1990). Mental models as contexts for interpreting texts. *J. Semantics 7*, 379–393.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition & Brain Theory 4*, 161–178.

Gentner, D. (1982). Why nouns are learned before verbs. In S.A. Kuczaj (ed.), *Language development* (Vol. 2, pp. 301-334). Erlbaum.

Gildea, P.M.; Miller, G.A.; & Wurtenberg, C.L. (1990). Contextual enrichment by videodisc. In D. Nix & R. Spiro (Eds.). *Cognition, education, and multimedia* (pp. 1-29). Hillsdale, NJ: Erlbaum.

Gillette, J.; Gleitman, H.; Gleitman, L.; & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition 73,* 135–176.

Gilman, E.W. (Ed.) (1989), *Webster's dictionary of English usage*. Springfield, MA: Merriam-Webster.

Goldfain, A. (2003). Computationally defining 'harbinger' via contextual vocabulary acquisition" [http://www.cse.buffalo.edu/~rapaport/CVA/Harbinger/ag33.CVA_harbinger.pdf].

Graesser, A.C., & Bower, G.H. (Eds.) (1990). *Inferences and text comprehension*. San Diego: Academic.

Granger, R.H. (1977). Foul-Up: A program that figures out meanings of words from context (pp. 172-178). In *Proc. 5th International Joint Conference on Artificial Intelligence*. Los Altos, CA: William Kaufmann.

Greene, A.J. (2010, Making connections. *Scientific American Mind 21(3)* (July/August), 22-29.

Guha, R.V. (1991). Contexts. Ph.D. dissertation. Stanford University.

Gunning, D.; Chandhri, V.K.; & Welty, C. (2010). Introduction to the special issue on question answering. *AI Magazine 31*(3), 11–12.

Haastrup, K. (1991). *Lexical inferencing procedures, or talking about words*.Tübingen: Gunter Narr.

Halliday, M.A.K. (1978). *Language as social semiotic*. Baltimore: Univ. Park Press.

Hansson, S.O. (1999). *A textbook of belief dynamics*. Dordrecht: Kluwer.

Harmon, J.M. (1999). Initial encounters with unfamiliar words in independent reading. *Research in the Teaching of English 33,* 304–318.

Harris, A.J., & Sipay, E.R. (1990). *How to increase reading ability, 9th Edition*. White Plains, NY: Longman.

Harsanyi, D. (2003, May 5). The old order. *National Review*, 45–46.

Hastings, P.M., & Lytinen, S.L. (1994a). The ups and downs of lexical acquisition. In *Proc. 12th National Conf. on AI* (pp. 754-759). Menlo Park, CA: AAAI/MIT Press.

Hastings, P.M., & Lytinen, S.L. (1994b). Objects, actions, nouns, and verbs. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 397-402). Hillsdale, NJ: Erlbaum.

Haugeland, J. (1985), *Artificial intelligence: The very idea*. Cambridge: MIT Press.

Higginbotham, J. (1985). On semantics. In E. Lepore (Ed.), *New directions in semantics* (pp. 1-54). Academic, 1987.

Higginbotham, J. (1989). Elucidations of meaning. *Ling. & Phil. 12*: 465–517.

Hillerman, T. (1990). *A thief of time*. Scranton, PA: HarperCollins.

Hirsch, Jr., E.D. (1987). *Cultural literacy*. Boston: Houghton Mifflin.

Hirsch, Jr., E.D. (2003). Reading comprehension requires knowledge—of words and of the world. *American Educator 27*(1): 10–13, 16–22, 28–29, 48.

Hirst, Graeme (2000). Context as a spurious concept. In *Proc., Conf. on Intelligent Text Processing & Computational Linguistics (Mexico City)*: 273–287.

Hobbs, J.R. (1990). *Literature and cognition*. Stanford, CA: CSLI.

Hobbs, J.; Stickel, M.; Appelt, D.; & Martin, P. (1993). Interpretation as abduction. *Artif. Intell. 63*: 69–142.

Horst, S. (2003). The computational theory of mind. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* [http: //plato.stanford.edu/entries/computational-mind/].

Hulstijn, J.H. (2003). Incidental and intentional learning. In C.J. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 349-381). Blackwell.

Ide, N.M., & Veronis, J. (Eds.) (1998). Special issue on word sense disambiguation, *Comput. Linguist. 24*(1).

Iwańska, L. & Zadrozny, W. (Eds.) (1997). Special issue on context in natural language processing, *Computational Intelligence 13*(3).

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford Univ. Press.

Jackendoff, R. (2006). Locating meaning in the mind (where it belongs). In R.J. Stainton (Ed.). *Contemporary debates in cognitive science* (pp. 219-236). Malden, MA: Blackwell.

Johnson, F.L. (2006). Dependency-directed reconsideration. PhD dissertation. Buffalo: SUNY Buffalo Dept. Computer Science & Engineering [http://www.cse.buffalo.edu/sneps/Bibliography/joh06a.pdf]

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge University Press.

Johnson-Laird, P.N. (1987). The mental representation of the meanings of words. *Cognition 25*, 189–211.

Jurafsky, D.; & Martin, J.H. (2008). *Speech and language processing, 2nd edition*. Upper Saddle River, NJ: Prentice Hall.

Kamp, H.; & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.

Kibby, M.W. (1980). Intersentential processes in reading comprehension. *J. Reading Behav. 12*, 299–312.

Kibby, M.W. (1995). The organization and teaching of things and the words that signify them. *Journal of Adolescent and Adult Literacy 39,* 208–223.

Kibby, M.W. (2007, Spring). Lean back and think. *.edu Alumni Newsletter of the UB Graduate School of Education*, pp. 13, 20 [http://gse.buffalo.edu/gsefiles/documents/alumni/GSE_newsltr_sp07.pdf].

Kibby, M.W.; Rapaport, W.J.; & Wieland, K.M. (2004). Helping students apply reasoning to gaining word meanings from context. Paper presented at the 49th Annual Convention of the International Reading Assn., Reno, NV.

Kibby, M.W.; Rapaport, W.J.; Wieland, K.W.; & Dechert, D.A. (2008). CSI: Contextual semantic investigation. In L. Baines (Ed.). *A teacher's guide to multisensory learning* (pp. 163-173). Alexandria, VA: Association for Supervision and Curriculum Development.

Kintsch, W. (1980). Semantic memory. In R.S. Nickerson (Ed.). *Attention and performance VIII* (pp. 595-620). Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension. *Psychological Review 95*(2): 163–182.

Kintsch, W. & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review 85*: 363–394.

Knuth, Donald. (1974). Computer science and its relation to mathematics. *Amer. Math. Monthly 81,* 323–343.

Kuhn, M.R., & Stahl, S.A. (1998). Teaching children to learn word meanings from context. *Journal of Literacy Research 30*: 119–138.

Lakoff, G. (1987). *Women, fire, and dangerous things*. University of Chicago Press.

Lammert, A. (2002). Defining adjectives through contextual vocabulary acquisition" [http://www.cse.buffalo.edu/~rapaport/CVA/Taciturn/lammert.full_report_ms.doc].

Langacker, R. (1999). *Foundations of cognitive grammar*. Stanford University Press.

Lenat, D.B. (1995). CYC: A large-scale investment in knowledge Infrastructure. *Commun. ACM 38*(11): 33–38.

Lenat, D.B. (1998). The dimensions of context-space. [http://www.cyc.com/doc/context-space.pdf].

Levin, B. (1993). *English verb classes and alternations*. University of Chicago Press.

Loui, M.C. (2000). Teaching students to dream. [https://netfiles.uiuc.edu/loui/www/gtc00.html].

Malory, Sir T. (1470). *Le morte darthur*. R.M. Lumiansky (Ed.). New York: Collier, 1982.

Martins, J.P. (1991). The truth, the whole truth, and nothing but the truth. *AI Magazine 11*(5), 7–25.

Martins, J., & Cravo, M. (1991). How to change your mind. *Noûs 25,* 537–551.

Martins, J.P., & Shapiro, S.C. (1988). A model for belief revision. *Artificial Intelligence 35*: 25–79.

McCarthy, J. (1993). Notes on formalizing context. In *Proc. 13th Int'l. Joint Conference on AI* (pp. 555-560). San Mateo, CA: Morgan Kaufmann.

McKeown, M.G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly 20*: 482–496.

Mervis, C.B.; & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology 32*: 89–115.

Meinong, A. (1904). Über Gegenstandstheorie. In R. Haller (Ed.), *Alexius Meinong Gesamtausgabe* II. pp. 481–535). Graz, Austria: Akademische Druck- u. Verlagsanstalt (1971).

Miller, George A. (1985). Dictionaries of the mind. In *Proc. 23rd Annual Meeting Assn. Comp. Ling.* (pp. 305-314). Morristown, NJ: Assn. for Computational Linguistics.

Miller, G.A. (1986), Dictionaries in the mind. *Language and Cognitive Processes 1*, 171–185.

Miller, G.A. (1991). *The science of words*. Scientific American Library.

Miller, G.A., & Gildea, P.M. (1985, June). How to misread a dictionary. *AILA Bulletin*: 13–26.

Minsky, M. (1974). A framework for representing knowledge. *Memo 306*.Cambridge, MA: MIT AI Lab.

Mondria, J.-A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics 12:* 249–267.

Mudiyanur, R. (2004). An attempt to derive the meaning of the word 'impasse' from context" [http://www.cse.buffalo.edu/~rapaport/CVA/Impasse/rashmi-seminar_report.pdf].

Mueser, A.M. (1984). *Practicing vocabulary in context: Teacher's guide*. Macmillan/McGraw-Hill.

Mueser, A.M., & Mueser, J.A. (1984). *Practicing vocabulary in context*.SRA/McGraw-Hill.

Murphy, J. (2000). *The power of your subconscious mind, revised edition*. Bantam.

Murray, K.M.E. (1977). *Caught in the web of words*. New Haven, CT: Yale.

Nagel, Thomas (1986). *The View from Nowhere*. New York: Oxford University Press.

Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly 19*, 304–330.

Nagy, W.E. & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge. In M.G. McKeown & M.E. Curtis (Eds.). *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ: Erlbaum.

Nagy, W.E.; Herman, P.A.; & Anderson, R.C. (1985). Learning words from context. *Read. Res. Quart. 20*, 233–253.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P., & Coady, J. (1988). Vocabulary and reading. In R. Carter & M. McCarthy (Eds.). *Vocabulary and language teaching* (pp. 97-110). London: Longman.

Ong, W.J. (1975). The writer's audience is always a fiction. *PMLA 90*(1), 9-21.

Parsons, Terence. (1980). *Nonexistent objects*. New Haven: Yale University Press

Pelletier, F.J. (2001). Did Frege believe Frege's principle? *Journal of Logic, Language and Information 10,* 87–114.

Piaget, J. (1952). *The origins of intelligence in children*. M. Cook (Trans.). New York: Int'l. Universities Press.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.). *Dimensions of mind* (pp. 148-179). New York: NYU Press.

Quillian, M.R. (1967). Word concepts. *Behavioral Science 12:* 410–430.

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quine, W.V.O. (1961). Two dogmas of empiricism. In W.V.O. Quine. *From a Logical Point of View, 2nd ed., rev*. (pp. 20-46). Harvard U. Press, 1980.

Rapaport, W.J. (1978). Meinongian theories and a Russellian paradox. *Noûs 12*: 153-180. Errata, *Noûs 13* (1979): 125.

Rapaport, W.J. (1981). How to make the world fit our language. *Grazer Philosophische Studien 14:* 1–21.

Rapaport, W.J. (1982). Unsolvable problems and philosophical progress. *Amer. Philosophical Qtly. 19,* 289–298.

Rapaport, W.J. (1986a). Searle's experiments with thought. *Philosophy of Science 53,* 271–279.

Rapaport, W.J. (1986b). Logical foundations for belief representation. *Cognitive Science 10,* 371–422.

Rapaport, W.J. (1991). Predication, fiction, and artificial intelligence. *Topoi 10,* 79–111.

Rapaport, W.J. (1998). How minds can be computational systems. *J. Exp. Theoret. Artif. Intell. 10,* 403–419.

Rapaport, W.J. (2000). How to pass a Turing test. *Journal of Logic, Language, and Information 9,* 467–490.

Rapaport, W.J. (2002). Holism, conceptual-role semantics, and syntactic semantics. *Minds and machines 12*, 3–59.

Rapaport, W.J. (2003a). What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. *Minds and machines 13,* 397–427.

Rapaport, W.J. (2003b). What is the 'context' for contextual vocabulary acquisition? In P.P. Slezak (Ed.), *Proc. 4th Intl. Conf. Cog. Sci./7th Australasian Soc. Cog. Sci. Conf.* (vol. 2, pp. 547-552). Sydney: U. New S. Wales.

Rapaport, W.J. (2005). In defense of contextual vocabulary acquisition. In A. Dey et al. (Eds.). *Modeling and using context* (pp. 396-409). Berlin: Springer-Verlag Lecture Notes in AI 3554.

Rapaport, W.J. (2005b). Castañeda, Hector-Neri. In J.R. Shook. Ed. *The dictionary of modern American philosophers, 1860-1960* (pp. 452-457). Bristol, UK: Theommes Press.

Rapaport, W.J. (2006a). The Turing test. In K. Brown (Ed.). *Encyclopedia of language and linguistics, 2nd ed.* (vol.13, pp. 151–159). Oxford: Elsevier.

Rapaport, W.J. (2006b). How Helen Keller used syntactic semantics to escape from a Chinese room. *Minds and machines 16,* 381–436.

Rapaport, W.J. (2009). Essential SNePS readings. [http://www.cse.buffalo.edu/~rapaport/snepsrdgs.html].

Rapaport, W.J. (compiler) (2010). A bibliography of theories of contextual vocabulary acquisition. [http://www.cse.buffalo.edu/~rapaport/refs-vocab.html].

Rapaport, W.J., & Ehrlich, K. (2000). A computational theory of vocabulary acquisition. In L. Iwańska & S.C. Shapiro (Eds.). *Natural language processing and knowledge representation* (pp. 347-375). Cambridge, MA: MIT Press. Errata: [http: //www.cse.buffalo.edu/~rapaport/Papers/krnlp.errata.pdf].

Rapaport, W.J.; & Kibby, M.W. (2007). Contextual vocabulary acquisition as computational philosophy and as philosophical computation. *Journal of Experimental and Theoretical Artificial Intelligence 19,* 1-17.

Rapaport, W.J.; & Shapiro, S.C. (1984). Quasi-indexical reference in propositional semantic networks. In *Proc. 10th Intl. Conf. Comp. Ling.* (pp. 65-70). Morristown, NJ: Assn. for Computational Linguistics.

Rapaport, W.J., & Shapiro, S.C. (1995). Cognition and fiction. In J.F. Duchan et al. (Eds.). *Deixis in narrative* (pp. 107-128). Hillsdale, NJ: Erlbaum.

Rapaport, W.J., & Shapiro, S.C. (1999). Cognition and fiction: An introduction. In A. Ram & K. Moorman (Eds.). *Understanding language understanding: Computational models of reading* (pp. 11-25). Cambridge, MA: MIT Press.

Rayner, K., & Well, A.D. (1996). Effects of contextual constraint in eye movements in reading. *Psychonomic Bulletin & Review* 3, 504–509.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *J. Experimental Psychology: General 118,* 219–235.

Resnik, M.D. (1967). The context principle in Frege's philosophy. *Phil. and Phenomenological Res. 27,* 356–365.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.). *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro et al. (Eds.). *Theoretical issues in reading comprehension*. Erlbaum.

Rumelhart, D.E. (1985). Toward an interactive model of reading. In H. Singer & R.B. Ruddell (Eds.). *Theoretical models and processes of reading, 3rd ed*. Newark, DE: International Reading Association.

Russell, B. (1905). On denoting. In B. Russell, *Logic and knowledge*. R.C. Marsh (ed.). (pp. 39-56). New York: Capricorn, 1956.

Russell, B. (1918). The philosophy of logical atomism. In B. Russell, *Logic and Knowledge*. R.C. Marsh (Ed.). (pp. 177-281). New York: Capricorn, 1956.

Saussure, F. de (1959). *Course in general linguistics*. C. Bally et al. (Eds.). W. Baskin (Trans.). New York: Philosophical Library.

Schagrin, M.L.; Rapaport, W.J.; & Dipert, R.R. (1985). *Logic: A computer Approach*. New York: McGraw-Hill.

Schank, R.C. (1982). *Reading and understandinge*. Hillsdale, NJ: Erlbaum.

Schank, R.C., & Rieger, C.J., III (1974). Inference and the computer understanding of natural language. *Artificial Intelligence 5,* 373–412.

Schatz, E.K., & Baldwin, R.S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly 21,* 439–453.

Schwartz, R.M. (1988). Learning to learn vocabulary in content area textbooks. *Journal of Reading*: 108–118.

Schwartz, R.M., & Raphael, T.E. (1985). Concept of definition. *Reading Teacher 39,* 198–203.

Schwartzmyer, N.P. (2004). Tatterdemalion: A case study in adjectival contextual vocabulary acquisition. [http://www.cse.buffalo.edu/~rapaport/CVA/Tatterdemalion/Tatterdemalion_final_report.pdf].

Seger, C.A. (1994). Implicit Learning. *Psychological Bulletin 115,* 163–196.

Shapiro, S.C. (1979). The SNePS semantic network processing system. In N. Findler (Ed.). *Associative networks* (pp. 179-203). New York: Academic.

Shapiro, S.C. (1982). Generalized augmented transition network grammars for generation from semantic networks. *American Journal of Computational Linguistics 8,* 12–25.

Shapiro, S.C. (1989). The CASSIE projects: An approach to natural language competence. In J.P. Martins & E.M. Morgado (Eds.). *4th Portuguese Conf. AI, Proc*. (pp. 362-380). Berlin: Springer-Verlag LNAI 390.

Shapiro, S.C. (1991). Cables, paths, and 'subconscious' reasoning in propositional semantic networks. In J.F. Sowa (Ed.). *Principles of semantic networks* (pp. 137-156). San Mateo, CA: Morgan Kaufmann.

Shapiro, S.C. (1992). Artificial intelligence. In S.C. Shapiro (Ed.), *Encyclopedia of artificial intelligence, 2nd ed*. (pp. 54-57). New York: Wiley.

Shapiro, S.C. (2001). Computer science: The study of procedures. [http://www.cse.buffalo.edu/~shapiro/Papers/whatiscs.pdf].

Shapiro, S.C. (2008). SNePS 2.7 user's manual. [http: //www.cse.buffalo.edu/sneps/Manuals/manual27.pdf].

Shapiro, S.C., & Rapaport, W.J. (1987). SNePS considered as a fully intensional propositional semantic network. In N. Cercone & G. McCalla (Eds.). *The knowledge frontier*. (pp. 262-315). New York: Springer-Verlag.

Shapiro, S.C., & Rapaport, W.J. (1991). Models and minds: Knowledge representation for natural-language competence. In R. Cummins & J. Pollock. Eds. *Philosophy and AI: Essays at the Interface*. (pp. 215-259)

Cambridge, MA: MIT Press.

Shapiro, S.C., & Rapaport, W.J. (1992). The SNePS family. *Computers & Math. with Applications 23,* 243–275.

Shapiro, S.C., & Rapaport, W.J. (1995). An introduction to a computational reader of narrative. In J.F. Duchan et al. (Eds.). *Deixis in narrative* (pp. 79-105). Hillsdale, NJ: Erlbaum.

Shapiro, S.C.; Martins, J.; & McKay, D. (1982), "Bi-Directional Inference", *Proc. 4th Annual Conf. Cog. Sci. Soc.*: 90-93.

Shapiro, S.C.; Rapaport, W.J.; Kandefer, M.W.; Johnson, F.L.; & Goldfain, A. (2007). Metacognition in SNePS. *AI Magazine 28,* 17–31.

Share, D.L. (1999). Phonological recoding and orthographic learning. *J. Exp. Child Psychol. 72,* 95–129.

Shieber, Stuart M. (Ed.) (2004). *The Turing test*. Cambridge, MA: MIT Press.

Simon, H.S. (1996). Computational theories of cognition. In W. O'Donohue & R.F. Kitchener (Eds.). *Philosophy of psychology* (pp. 160-172). Sage.

Simon, H.S.; & Newell, A. (1962). Simulation of human thinking. In M. Greenburger (Ed.). *Computers and the world of the future* (pp. 94-114). Cambridge, MA: MIT Press.

Sinclair, J. (Ed.) (1987). *Collins COBUILD English Language Dictionary*. London: Collins.

Singer, M.; Revlin, R.; & Halldorson, M. (1990). Bridging-inferences and enthymemes. In A.C. Graesser & G.H. Bower (Eds.). *Inferences and Text Comprehension* (pp. 35-51). San Diego: Academic.

Srihari, S.N.; Collins, J.; Srihari, R.; Srinivasan, H.; Shetty, S.; Brutt-Griffler, J. (2008). Automatic scoring of short handwritten essays in reading comprehension tests. *Artificial Intelligence 172,* 300–324.

St.-Exupéry, A. de (1948). *Wisdom of the sands*, cited by M. Sims, *Chron. Higher Ed.* (15 August 2003): B4.

Stalnaker, R.C. (1999). *Context and content*. New York: Oxford University Press.

Stanovich, K.E. (1986). Matthew effects in reading. *Reading Research Quarterly 21,* 360–407.

Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.). *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.

Sternberg, R.J., & Powell, J.S. (1983). Comprehending verbal comprehension. *American Psychologist 38:* 878–893.

Sternberg, R.J.; Powell, J.S.; & Kaye, D.B. (1983). Teaching vocabulary-building skills: A contextual approach. In A.C. Wilkinson (Ed.). *Classroom computers and cognitive science* (pp. 121-143). New York: Academic.

Storkel, Holly L. (2009), "Developmental Differences in the Effects of Phonological, Lexical and Semantic Variables on Word Learning by Infants", *Journal of Child Language* 36: 291–321.

Suh, S., & Trabasso, T. (1993). Inferences during reading. *Journal of Memory and Language 32,* 279–300.

Swanborn, M.S.L., & De Glopper, K. (1999). Incidental word learning while reading. *Rev. Educ. Res. 69,* 261–285.

Szabó, Z.G. (2007). Compositionality. In E.N. Zalta (Ed.). *Stanford Encyclopedia of Philosophy* [http://plato.stanford.edu/archives/spr2007/entries/compositionality/].

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Tayler, J. (1997). Transylvania today. *Atlantic Monthly 279*(6): 50–54.

Taylor, W.L. (1953). Cloze procedure. *Journalism Quarterly 30,* 415–433.

Thagard, P. (1984). Computer programs as psychological theories. In O. Neumaier (Ed.). *Mind, language, and society* (pp. 77-84). Vienna: Conceptus-Studien.

Turing, A. (1950). Computing machinery and intelligence. *Mind 59,* 433-460.

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.

Waite, M. (Ed.) (1998). *Little Oxford dictionary, rev. 7th ed*. Oxford U. Press.

Webster, M.A.; Werner, J.S.; & Field, D.J. (2005). Adaptation and the phenomenology of perception. In C.W.G. Clifford & G. Rhodes (Eds.), *Fitting the Mind to the World* (pp. 241-277). Oxford: Oxford Univ. Press

Werner, H., & Kaplan, E. (1952). The acquisition of word meanings. *Monographs, Society for Research in Child Development* 15 (1950) 51(1).

Wesche, M., & Paribakht, T.S. (Eds.) (1999). Incidental L2 vocabulary acquisition. *Stud. Second Lang. Acq. 21*(2).

Widdowson, H.G. (2004). *Text, context, pretext*. Malden, MA: Blackwell.

Wieland, K.M. (2008). Toward a unified theory of contextual vocabulary acquisition. PhD dissertation. Buffalo: SUNY Buffalo Department of Learning & Instruction.

Wikipedia (2007). Phalacrosis. [http://en.wikipedia.org/w/index.php?title=Phalacrosis&oldid=171720826].

Wittgenstein, L. (1958). *Blue and brown books, 2nd ed*. Oxford: Blackwell.

Wolfe, A. (2003, May 30). Invented names, hidden distortions in social science. *Chronicle Higher Ed*: B13–B14.

Woods, W.A. (1970). Transition network grammars for natural language analysis. *Comm. ACM 13:* 591–606.

Zalta, Edward N. (1983). *Abstract objects*. Dordrecht: D. Reidel.

Sentences of this form ...:	**... are encoded as a SNePS network representing this proposition:**
x is *P* ; i.e., NP is Adj; e.g., "Fido is brown"	*x* is an object with property *P*[38]
x is a *P*; i.e., NP_{indiv} is an NP_{common}[39] e.g., "Fido is a dog."	*x* is a member of the class *P*
a *P* is a *Q*; i.e., An NP_{common} is an NP_{common} e.g., "A dog is an animal."	*P* is a subclass of the superclass *Q*
x is *y*'s *R*; i.e., NP is NP's NP e.g., "This is Fido's collar."	*x* is an object that stands in the *R* relation to possessor *y*[40]
x does *A* (with respect to *z*) e.g., Fred reads (a book)	agent *x* performs the act of: doing action *A* (with respect to object *z*)
x stands in relation *R* to *y* e.g., Fido is smaller than Dumbo	relation *R* holds between first object *x* and second object *y*
A causes *B*	*A* is the cause of effect *B*
x is a part of *y*	*x* is a part of whole *y*
x is a *PQ* e.g., "Fido is a brown dog."	*x* is a member of the class *P* & *x* is a member of the class *Q*
x is a *PQ* e.g., "This is a toy gun." (cf. §6.1), "This is a small elephant." "This is a fire hydrant."	*x* is a member of the class whose class modifier is *P* and whose class head is *Q*
x is (extensionally the same as) *y* e.g., "Superman is Clark Kent."	*x* and *y* are equivalent[41]
x is a synonym of *y*	*x* and *y* are synonyms

Table 1: Basic SNePS Propositions

[38]More precisely, "*x* is an object with property *P*" is represented by a network of the form: The English word *x* expresses an object with a property expressed by the English word *P*.
[39]I.e., a sentence consisting of a noun phrase representing an individual, followed by 'is a', followed by a common-noun phrase.
[40]For more information on the possessive "*x* is *y*'s *R*", see Rapaport 2006b.
[41]Shapiro & Rapaport 1987.

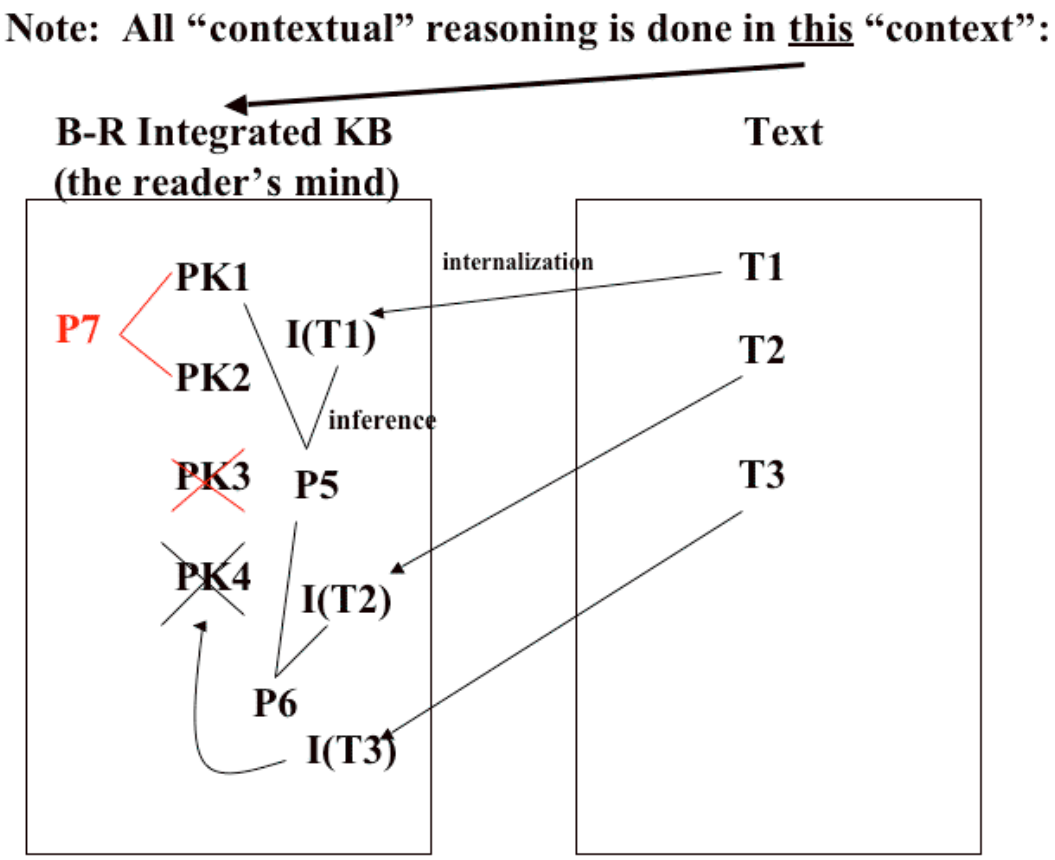

Figure 1: A belief-revised, inegrated knowledge base and a text