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

September 2, 2009

Abstract

Deliberate contextual vocabulary acquisition (CVA) is a reader's ability to figure out *a* (not *the*) meaning of 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.

1 A Computational Theory of Contextual Vocabulary Acquisition

What should a reader do when confronted with an unfamiliar word?

The usual response is 'Look it up in the dictionary!'. But, all too often, readers don't do this, or don't want to. Sometimes, readers don't even 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: 53), and words that are highly likely to be chosen in a cloze task are skipped more than others (Rayner & Well 1996: 504). Paradoxically, "children may misread or ignore unfamiliar words *without* jeopardizing comprehension" (Bowey & Muller 2005: 205, citing Share 1999; our emphasis).

Other times, readers realize that they don't understand a word but are too lazy (or too embarrassed) to do anything about it. Perhaps they hope that the word wasn't all that important and won't be used again. Or perhaps they are discouraged by previous, unsuccessful attempts to look words up.

Or perhaps not: They might realize that the word is unfamiliar (or vaguely familiar; or familiar, but being used unfamiliarly). And they might be curious as to what it might mean, i.e., what the author had in mind when using it. So they can ask someone. If no one's 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 there's no dictionary handy? Or the only one handy doesn't contain the word? Or it does, but the definition doesn't seem appropriate for the context in which it is being used? Or the definition is convoluted or otherwise unhelpful (cf. Rapaport & Kibby 2007, §2.1)? With the exception of learner's dictionaries designed primarily for ESL audiences, dictionaries are notoriously difficult to use and their definitions notoriously difficult to interpret.[1]

So, what *can* readers do? They 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 enables them to feed themselves for a lifetime, so *giving* a reader a definition tells them what one word means, but *teaching* them CVA enables them to become better readers.

We make four principal claims: (1) "Texts . . . are . . . the soil in which word-meaning understandings are grown" (Wieland 2008); the reader's prior knowledge is the water that enables that growth. (2) A procedure for successful CVA can be expressed in terms so precise that they can be programmed into a computer. (3) That computational procedure can then be converted into a curriculum teachable to human readers. (4) This CVA procedure can help readers improve both their vocabulary and their reading comprehension.

But how does one do CVA? And what is the 'context' that new vocabulary is acquired in? Readers do CVA 'incidentally', i.e., unconsciously (§2.2). They can also do it 'deliberately', i.e., consciously or intentionally (§2.3; cf. Hulstijn 2003). And they can be taught how to do it. But how well are they taught it? And *where* are they taught it: in an isolated unit on vocabulary? a unit on reading comprehension? a literature class? a second-language class? 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?

No computer programmed to process natural language can yet convince a native speaker of English that it 'understands' English (Turing 1950; cf. Shieber 2004, Rapaport 2006a). But there are many programs that process English text—that can parse a text, construct semantic interpretations of it, answer simple questions about it, or do information retrieval. (For a survey of the state of the art in computational natural-language processing, see Jurafsky & Martin 2008.) In a future 'golden age' of cognitive science and artificial intelligence (AI), computers might be able to fully converse with humans.

---

[1] Examples are myriad (see Miller 1985, 1986; Miller & Gildea 1985). Our personal favorites are: *Webster's Ninth New Collegiate Dictionary* (Mish 1983: 259) defines 'college' as "a body of clergy living together and supported by a foundation". *Merriam-Webster OnLine* [http://www.m-w.com/dictionary/] leads the reader from 'infract' through 'infringe' to 'encroach', then defines 'encroach' as 'to enter by gradual steps or by stealth into the possessions or rights of another'. And even the *Oxford English Dictionary Online* [http://dictionary.oed.com/] defines 'maze' as 'labyrinth', and 'labyrinth' as 'maze' (along with more helpful information).

What should such a computer do when confronted with an unfamiliar word? It could, of course, 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' a meaning for the unfamiliar word from 'context'. Here, 'to figure out' would mean 'to compute'. Can a computer compute a meaning for an unfamiliar word from context? Under certain (reasonable) circumstances, it can. Moreover, we can adapt its methods for doing this for teaching human readers to compute meanings in the same way.

This seems to reverse the usual way in which computers are programmed (or 'taught') to do certain cognitive tasks. According to 'good old-fashioned', symbolic AI (Haugeland 1985), if we know how to explicitly teach some 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. And, if we don't know how to teach some task—how would you teach a human to see?—then it's very hard (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 includes reading educators as well as a philosopher of language and logic who is also an AI researcher. Our project has twin goals: (1) to develop a computational, cognitive theory of deliberate CVA, and (2) to adapt our computational CVA strategies to an educational curriculum for teaching them to students. We have 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. But it is a two-way street: We are also improving our program on the basis of observations of good readers doing CVA (Wieland 2008). To this end, we have been studying the CVA literature from a variety of disciplines, including computational linguistics, psychology, first- and second-language (L1, L2) acquisition, and reading education. Generally speaking, each discipline tends to ignore each other's literature. (For surveys, see Rapaport & Ehrlich 2000, Wieland 2008; Rapaport 2009 is a (partial) bibliography of CVA research in all of these fields.)

2 The Nature of Contextual Vocabulary Acquisition

2.1 Preliminary Terminology

'CVA' is our term for "acquisition of word meaning from context", "word-meaning derivation", or "lexical inferencing". We use it to 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.

For example, consider this passage containing the word 'tatterdemalion' (here assumed to be unfamiliar):

> 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.)

What might 'tatterdemalion' mean? One informant thought out loud as follows:

> 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 .... 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, it appears that she understands the word as connoting something roughly second-class.

Our computational system reasoned roughly as follows (see Schwartzmyer 2004):[2]

---

[2]What follows is a translation of the computer transcript into full English sentences.

Comfortable and tatterdemalion are in a "but" relationship. If object 1 is in a "but" relationship with object 2, and object 2 is a positive attribute, then object 1 is a negative attribute, equivalent to a negative quality. Comfortable is a positive attribute; therefore, 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 two dollars, then they are not expensive, which means that 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.

Although CVA can be used in ordinary conversation, watching TV, etc., we focus only on CVA during reading. Everything we say should carry over to spoken domains, and there have been some studies of oral CVA (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!).

We first need to clarify some terminology. Interdisciplinary cognitive scientists, especially, face the problem that many terms are used differently by different researchers, without any notice of the differences, often resulting in confusion. One should always try to figure out (from context, if by no other way!) how an author uses such terms. On the other hand, one should never *use* any of these terms without clearly explaining how one is using it. The words 'definition', 'sense', and 'meaning' are among them.

(1) 'Definition' is a loaded term; it brings to mind the dictionary-style of definition, which is *not* what we're after. In Ehrlich 1995 and our earlier papers on CVA, we did talk about the end-product of CVA being a "dictionary-like definition". But the emphasis was always intended to be on '-like': We wanted the output of our computer programs to be expressed in the sort of language that dictionaries use, e.g., "A brachet is a dog that hunts". We certainly didn't mean that the output would be a complete, correct, and highly detailed "definition" (in the bad sense of that word; see note 1).

(2) 'Sense' has different meanings in different disciplines. For many people, it is just a synonym for 'meaning' without the bad overtones of 'definition'; for philosophers, 'sense' is a technical term.

In 1892, the German philosopher Gottlob Frege analyzed meaning into two components. In German, he called these '*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 (a word, phrase, or sentence) referred to, or denoted. For instance, '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. Yes, there are pictures and stories about them, but (unfortunately) they don't exist.

Does that mean that 'unicorn' is meaningless? No: It has a sense (a *Sinn*), even though it doesn't have a referent. But 'snow' has a referent *in addition to* a sense.[3] In fact, the sense of 'snow' *determines* its referent: 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) identify a real unicorn if you ever saw one (even though you won't).[4] So, every word has a sense, but not all words have referents.

---

[3] Strictly speaking, for Frege, a linguistic expression doesn't have a *Bedeutung*. Rather, every linguistic expression ('snow', 'unicorn') has a "sense" (*Sinn*), and some, but not all, *senses* (*Sinne*) have "referents" (*Bedeutungen*); e.g., the sense of 'snow' has a referent (the cold, wet, white stuff), but the sense of 'unicorn' doesn't. Furthermore, multiple "senses" (*Sinne*) can have the same "referent" (*Bedeutung*); e.g., the senses of the expressions 'the morning star' and 'the evening star' have the same referent: Venus.

[4] It has been said that if we saw a creature that satisfied the definition of 'unicorn', we might refuse to apply that term to it in a literal way, on the grounds that we antecedently know that unicorns don't exist. So, if we saw a unicorn-like creature, we might call it "unicorn-like" but wouldn't say that it was the celebrated creature of myth (see Associated Press 2008).

Fregean senses are "objective"; each unambiguous word has just one sense, which, somehow, all minds "grasp". Each of us also has our own concepts attached to words, psychologically unique to each of us. Frege disdained these as unscientific; he wanted to de-psychologize meaning. The meanings computed by our programs are "psychological" in this way, hence similar to, but not exactly the same as, Fregean senses. Because 'sense' is a term of art for philosophers, linguists, and computer scientists, we avoid it when possible.

(3) 'Meaning' might well be the most neutral term covering all this. But we should not talk about "the" meaning of a word. A word has *many* meanings. Not only are most words ambiguous (or polysemous), but each of us has our "own" (psychological, in Frege's terminology) meanings for words. This is the focus of the semantic theories of "cognitive linguists" (Langacker 1999, Talmy 2000, Jackendoff 2002). There are problems about how we can understand each other if we all mean something different by our words, but they can be overcome (see §4.1.3 and Rapaport 2003a).

We prefer 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 the meaning that is produced by CVA is a *hypothesis* that is constructed and assigned to the word by the reader, rather than being "the correct" (dictionary-style) definition that, in some sense, "belongs" to the word:

> [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.)

But 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, Ch. 6), as Wittgenstein's last sentence suggests. Similar ideas have been put forth by contemporary cognitive scientists: 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 (2006: 38; in press) 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 an idea suggested by Dave Rumelhart in the late 1970s, 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. These effects are lawful and predictable, and it is these regularities that we intuitively take as evidence of word knowledge." (Cf.: "Words don't *have* meaning; they're *cues* to meaning!"; Elman, oral presentation, 29th Annual Conf., Cogn. Sci. Soc.; Nashville, 3 August 2007, based on Elman 2007.)

CVA is neither a process of discovery of a word's "correct" meaning (whatever that might be) nor (necessarily) determining *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 the hypothesis when future occurrences of the word are encountered. The reader only has to determine *a* meaning for the word (as it appears in the text) *that enables the reader to understand 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.

(4) Almost everyone working on this topic believes that it is possible to "figure out" a meaning for a word "from context". Other terms for this process include 'construct', 'deduce', 'derive', 'educe', 'guess', 'infer', and 'predict'. Because the CVA process of figuring out a meaning for a word from context 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 theory of mind—human readers do. (On the computational theory of mind, see: Putnam 1960, Fodor 1975, Rapaport 1998, Horst 2003.) But we prefer 'figure out' to any of the other terms

that appear in the literature, such as 'deduce' (which is too narrow) or 'infer'. Herbert Simon (1996: 171) 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 the same coin. We especially dislike "guess a meaning", with its connotation of randomness, as well as for its lack of guidance for readers doing CVA (see §§2.3.2, 7.4, below). However, because the phrase "the reader computes a meaning" is awkward at best, we shall use "figure out". (In any case, "figure out" may be metaphorical for "compute"!)

2.2 Incidental CVA

CVA is not restricted to fluent readers faced with a new word, incrementally increasing their vocabulary. Rather, 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). *Excellent* students who read a great deal may know 120,000 word families (Miller 1991). But learning even the low-end estimate of 45,000 words by age 18 means learning an average of some 2500 words each year; yet no more than 400 words per year are directly taught by teachers (Nagy & Anderson 1984)—4800 words in 12 years of school. Therefore, around 90% of the words we know and understand must have been learned from oral or written context. Learning words from context 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; see Christ 2007 for a recent study of incidental CVA from our research group's perspective), and it is very likely the result of unconscious inductive inference (Reber 1989, Seger 1994).

2.3 Deliberate CVA

But some CVA is the result of "deliberate", or conscious and active, processes of hypothesizing a meaning for unfamiliar words from context. How do readers do this? The psychology, L1, and L2 literatures suggest various strategies, most of which are quite vague (e.g., 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). How is deliberate CVA taught in classroom settings? Not well.

2.3.1 Mueser & Mueser

Mueser & Mueser (1984; Mueser 1984) have a workbook (used at one time 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.

2.3.2 Nation et al.

Clarke & Nation (1980: 212) offer the following strategy:

step 1: "... look at the word itself and its surroundings to decide on the part of speech." This is fairly straightforward, depending only on the reader's knowledge of grammar. Note that the reader's knowledge of grammar is not in the text, but "in" the reader's mind (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: 67).

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 causal, temporal, class-membership information, etc., of the sort recommended by Sternberg et al. 1983, Sternberg 1987, and the others cited above (§2.3).

step 4: "... *guess... the word* and check... that the guess is correct" (our emphasis).

This is a bit more useful in real life than the Mueser technique. But put yourself in the position of a poor reader forced to 'guess'. What does she or he really have to do? How should the reader put all the data gathered in steps 1–3 together in order to produce a 'guess'? And if the reader was supposed to guess in the first place, why make him or her go through that data-gathering procedure?

In particular, all the important work that we are concerned with is hidden in the first part of step 4, which is reminiscent of a famous Sidney Harris cartoon[5] showing a complicated mathematical formula, in the middle of which appears the phrase, 'then a miracle occurs'! The cartoon shows a man looking at this and commenting to the man who wrote it: "I think you should be more explicit here ...".

To be fair, Clark and Nation point out that by 'guess' they really mean 'infer' (Clark & Nation 1980: 211, footnote 1), but they still haven't offered the reader any guidelines for *how* to make the inference.[6]

Later on, they say (our emphasis):

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

This is supposed to provide 'detail'. Did we miss something between 'guess the meaning and then check his guess" and "There are three ways of checking"? Nation & Coady (1988: 104–105) offer "an elaboration of" Clarke & Nation's strategy. Their "elaboration" has *five* steps (cf. Nation 2001: 257):

1. Finding the part of speech of the unknown word.
2. Looking at the immediate context of the unknown word and simplifying this context if necessary.
3. Looking at the wider context of the unknown word. This means looking at the relationship of the clause containing the unknown word and surrounding clauses and sentences.
4. Guessing the meaning of the unknown word.
5. Checking that the guess is correct.

---

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

[6] Mondria & Wit-de Boer (1991: 251–252) say explicitly that to guess is to infer from context. And computer scientist Michael Loui 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". (Loui 2000.)

Splitting one conjoined step into two single steps is hardly 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 merely 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.* (Cf. Schagrin et al. 1985: xiii.) This is what we do in the curriculum that we present in §7, below.

2.3.3 Sternberg et al.

Psychologist Robert Sternberg and colleagues (Sternberg & Powell 1983; Sternberg, Powell, & Kaye 1983; Sternberg 1987) call CVA 'learning from context'. But there are really two different notions: CVA is, roughly, figuring out a meaning for a word solely from its textual context plus the reader's background information (we will be more precise about this in §3, below). The author assumes that the reader already knows a meaning for each word in the text and, therefore, does not necessarily construct the text so as to purposefully provide information for learning a meaning for an unfamiliar word in it.

This must be distinguished from learning word meanings from purposeful contexts: teaching a meaning for a word by presenting the word in a (sentential) context designed for this purpose, i.e., 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 'learning from context' in the sense of CVA, but between these and learning word meanings from purposeful contexts. Most of the experiments that Sternberg cites probably considered learning from context designed for teaching, not from contexts not necessarily so designed, as is the case with CVA. (The rest of Sternberg's paper is about CVA as we do it, so to some extent these comments might be beside the point, 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 our discussion of Beck et al. 1983, §4.1, below; cf. footnote 34, below.)

However, these 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: 91.)

This passage contains *two* unknown words, meanings for both of which are easily figured out. Sternberg et al. 1983: 131–133 (cf. Sternberg 1987: 92–94) 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 ever shown this 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

8

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). In fact, it means 'baldness'. [7]

But suppose a reader decided it meant 'grey hair'. Has that reader figured out an 'incorrect' meaning? Technically, yes; but does it really matter? Such readers would have figured out *a* (reasonable) meaning that enable them to understand the passage and to 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.

All of this aside, Sternberg's goals are the same as ours: to produce a theory of how to do and teach CVA. But his theory (in Sternberg et al. 1983) is quite vague. Yes, we are given eight cues and seven mediating factors (one of which, background knowledge, plays a much larger role in our theory than in his), but they are devoid of detail and contain no instructions on what to do with them (except for an example or two). Compounding this is Sternberg et al.'s 'general strategy for context use' (i.e., for doing CVA):

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 of the unfamiliar word from the general context that *follows* the word ….";

step 3: "Attempt to infer the meaning of the unknown word by looking at the word parts …" (i.e., by 'looking at' its morphology); [8]

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 of the unknown word 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 …." (Sternberg et al. 1988: 139–141; our emphases.)

---

[7] In 2003, WJR attempted to find this in a dictionary. At that time, it was hard to find. By 2007, a Google search turned up several websites with definitions, the most curious of which was *Wikipedia* (2007), which (erroneously) calls it a 'nonce' word, citing one of WJR's websites about precisely this sentence [http://www.cse.buffalo.edu/∼rapaport/CVA/acapnotic.html]!

[8] Sternberg calls the use of morphological information 'internal' context (as opposed to the 'external' context consisting of the surrounding 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). And, as Sternberg has noted, internal context can't be used in isolation from external context (Sternberg & Powell 1983: 886; Sternberg et al. 1983: 135, 138). For example, 'inflammable', on the face of it, ought to mean 'not flammable', because it has 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, we would simply have to have a procedure in the algorithm 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, an interesting project for the future is to have readers use CVA to learn the meanings of *affixes* from context! (We do incorporate this into our curriculum; see §7.3, below.)

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. Also, step 5 appears at best to be part of the needed specifications for steps 1 and 2, and at worst appears to merely repeat them. In any case, step 5 is no more detailed than Clarke & Nation's steps 3 or 4. Many questions remain: *What* should the reader do with the information found in steps 1–5? *How* should the reader make the required inference? Is it as simple as a deductive inference? And *how* should the reader "construct" a definition (step 6)? [9]

2.3.4 The Computational Approach to CVA

Part of the problem is that, while many authors suggest what contextual clues to look for, few (if any) provide specific advice on what to *do* with these clues once they have been found. (1) How should global and local text comprehension be employed? (2) What reasoning and other cognitive processes are useful? And (3) how should prior knowledge be applied?

Previous views of CVA privilege concrete textual clues (such as definitions, synonyms, antonyms, examples, apposition, comparison, contrast, etc.; see Ames 1966, Baumann et al. 2003). We privilege the reader's comprehension, prior knowledge, and thinking. The reader brings at least as much power and information to the CVA process as the text provides in terms of contextual or morphological clues. As Baumann et al. 2003: 485 say, '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." Or else readers' understanding is a function of how much *textual analysis* (of the sort involved in our CVA procedures) is involved, as opposed to being a function of merely *looking* for clues that are either obvious or possibly not even there.

One other suggested strategy is *slightly* more explicit (perhaps): 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, cf. Denning 2007). What readers need to be taught is 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 reasoning within a context 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 larger vocabularies and better reading comprehension.

Learning concepts and their words—especially when the concept is not part of the reader's prior knowledge and 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: 210). 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"

---

[9]On the other hand, Sternberg 1987 seems to go into considerably more detail (pp. 90–94, 99–100). But the detail is not presented in an orderly way that could be systematically applied by a reader; it is a grab-bag of techniques, not a procedure.

[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 in a painting, but even an unsophisticated reader has the prior knowledge necessary to learn this "pentimento" 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 either, but they would have to be relatively knowledgeable about statistics to learn it. (See Kibby 1995.)

(Piaget 1952) or "integrating" (Kintsch 1988) or "consolidating" (Hansson 1999) the newly learned concept with prior knowledge, which might include inference, belief revision, or reorganizing existing cognitive schemata.

To spell out all the steps of inference (as we sketched for 'phalacrosis') requires a detailed sequence of steps (which, to be fair, would no doubt have exceeded the page limits Sternberg was given by his editors). These detailed steps constitute 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 cognitive scientists like Allen Newell and Herbert Simon meant when they said that such computer programs *are* theories of psychological behavior (Newell et al. 1958: 151–153). We are investigating ways to facilitate readers' natural CVA by developing a rigorous computational theory of how context is used and by 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 (i.e., written as a computer program) (initially by Ehrlich 1995) in a propositional semantic-network knowledge-representation-and-reasoning system (SNePS; Shapiro & Rapaport 1987, 1992, 1995; Shapiro & Group 2007). (A semantic network is a graphical structure for representing relationships among concepts, much like what reading educators have termed "concept maps". A knowledge-representation-and-reasoning "system" is a computer program that represents and reasons about information. Computational CVA systems have also been developed by Granger 1977, Haas & Hendrix 1983, Berwick 1983, Zernik & Dyer 1987, Hastings & Lytinen 1994ab, Siskind 1996, etc.; see Rapaport & Ehrlich 2000, §8, for our evaluation of these.) Our computational system begins with a stored knowledge base containing SNePS representations of relevant prior knowledge. It takes as input SNePS representations of a passage containing an unfamiliar word. The processing begins with inferences drawn from these two, integrated sources of information.[11] When asked to define the word, it applies definition algorithms (for nouns and for verbs; adjectives and adverbs are under investigation) that deductively search the resulting 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, 2004; Rapaport & Ehrlich 2000; Rapaport & Kibby 2007; Rapaport 2003b, 2005; Shapiro et al. 2007. We are investigating ways to make our system more robust and to embed it in a natural-language-processing system.)

And what a curriculum should do is teach such an algorithm (minus any implementation-dependent details) to students. To do this, the algorithm 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"

'Context' is another notoriously vague term. It is likely that no two researchers mean the same thing by it. But it is essential for our purposes that we work with a reasonably precise characterization of it.

---

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

3.1 Word and Context

We begin by considering the relationship between a word and its surrounding *textual* context. The smallest surrounding textual context of a word would probably be a phrase of which the word is a grammatical constituent. Typically, one thinks of a word's textual context as the *sentence* in which the word occurs.

Which comes first: the meaning of a sentence, or the meaning of a word? Most people would probably opt for the latter:

(1) The meanings of words are primary, and the meaning of a sentence depends on the meanings of its constituent words.

This not only seems obvious, but underlies the standard compositional (or 'recursive') view of semantics espoused by most contemporary linguists (for discussion, see 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: 800, in the entry 'reason is because", in which it is argued that, in the context 'the reason is because", 'because' means 'that"). Consequently, some researchers hold that

(2) the meanings of sentences are primary, and the meaning of a word depends on the meanings of the sentences that it occurs in.

Although Frege is well known among philosophers for espousing (1), at one time he also held (2):

> Only in the context [*Zusammenhange*] of a sentence [*Satz*][12] do words mean [*bedeuten*][13] something. (WJR's translation of Frege 1884, §62; cf. §60.)[14]

The 20th-century American philosopher Willard Van Orman Quine (1961: 39) urged something similar: 'the primary vehicle of meaning came to be seen no longer in the term but in the statement." [15]   And Bertrand Russell's (1905) theory of definite descriptions can be understood this way, too: One can't say what 'the' means; one can only say 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* the meaning of (or: a meaning for) 'the'. This has also been recognized in the field of vocabulary acquisition in the teaching of reading: Deighton (1959: 2) said, 'while context always *determines* the meaning of a word, it does not necessarily *reveal* that meaning."

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.[16] We return to this later (§§3.5, 4.1.5.1, 7).

---

[12]The Austin translation has 'proposition', but, in this context, '*Satz*' has to mean 'sentence", because Frege is talking about words, not concepts.

[13]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.

[14]For further discussion of this principle of Frege's, see Resnik 1967, Pelletier 2001.

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

[16]'[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: 413).

Beck, McKeown, & Omanson (1984: 4) 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, the meaning of a word can depend on its context in the sense that (1) the intended meaning 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 or extend—in short, change—the meaning of a word.

3.2 The Nature of Context for CVA

But what is 'context'? Most CVA researchers in all disciplines have in mind *written* contexts, as opposed to *spoken* contexts and as opposed to a broader notion of 'context' that might include visual or 'situative' information (speaker, location, time, etc.). Still, there is ambiguity (Engelbart & Theuerkauf 1999): Informally, many researchers say something like this: 'The reader can infer/guess/figure out, etc., the meaning of a word *from context* ...' (e.g., Werner & Kaplan 1952, McKeown 1985, Schatz & Baldwin 1986). Sometimes they *say* that, but *mean* something like this: '... from context *and* the reader's background knowledge' (e.g., 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' (e.g., Nation & Coady 1988; see also 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*.[17] What exactly is—or should be—meant by *contextual* vocabulary acquisition?

Graeme 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 (e.g., McCarthy 1993) that take the term as primitive. He 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.

A clue to the nature of context as needed for our purposes can be found in our CVA software: We represent, in a *single* semantic-network knowledge base, *both* the information in the text *and* the reader's background knowledge (Rapaport & Ehrlich 2000, and §3.4 and Fig. 1, below). This strongly suggests that the relevant 'context' for CVA of the unknown word is (at least a subpart of) the entire surrounding *network*.

What follows is a sequence of terms and their definitions, leading to a proposal for the proper definition of 'context' as it should be used in CVA and that is consistent with our computational cognitive approach.

3.3 Preliminary Definitions

We will use the expression **unfamiliar term for a reader** to mean a word or phrase that the reader either has never seen before, or has only the vaguest idea about its meaning, or that is being used in a new or unfamiliar way. (Kibby 1995 discusses levels of familiarity.) For convenience, let's symbolize this unfamiliar term as 'X'.

A **text** will be a (written) passage. It could be as short as a sentence or as long as an entire book (in a novel, knowledge of characters, settings, and themes might all be needed for CVA), and it will usually contain X. It is not essential that the text be written: Presumably the same techniques could be applied to

---

[17]Cf. our comment in §2.3.2 on Clarke & Nation's step 1.

oral CVA (though there would be attentional and memory limitations; see §§2.1, 7); in any case, most CVA research concerns texts that are read, rather than heard.

The next definition uses a possibly awkward term of art, but it serves a useful role, and others have used it before: The **co-text of** X **as it occurs in** some text T is the entire text T "minus" X (i.e., the entire text surrounding X; Catford 1965: 31n2.)[18] So, if X = 'brachet', and our text is:

(T1) There came a white hart[19] 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 X 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.

The underscore marks the location of the missing X. 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, however, the reader is not usually trying to *find* a known-but-missing *word* (a "binary" task at which one either succeeds or else fails). Rather, the reader is hypothesizing a *meaning* for a visible-but-unknown word (a "continuous" task at which one can do well or poorly or anywhere in between).[20]

The **reader's prior knowledge** is the "knowledge" that the reader has when s/he *begins* to read the text *and* that s/he is able to recall as needed while reading. Plato analyzed knowledge as justified, *true* belief (*Theaetetus* 201). Because some of what readers think they know is likely mistaken, 'belief' is a more appropriate word than 'knowledge'. But we can use both, as long as it's clear that prior "knowledge" need not be true. Similar terms are used by other researchers; however, they all have slightly different connotations:

1. "*Prior* knowledge" suggests knowledge that the reader has *before* reading, i.e., the beliefs that the reader brings to the text and has 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.

2. "*Background* knowledge" lacks that temporal connotation, but is otherwise synonymous for our purposes. It might, however, more usefully refer to the information that the text's *author* assumes that the reader should have (cf. 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 [the] ... words [in the text] evoking somewhat the same associations in the

---

[18] The term seems to originate with Catford. Halliday 1978: 133 cites Catford, and Brown & Yule 1983: 46–50 cite Halliday; cf. Haastrup 1991, Widdowson 2004. *Pace* Schatz & Baldwin 1986, the co-text should not be *limited* to a 3-sentence window around X (§4.2.2.3, below).

[19] We are assuming that the reader knows what a hart is but not what a brachet is, so that the reader can use his or her knowledge of harts to help define 'brachet'. If the reader doesn't know what a hart is, then he or she would have to try to figure out a meaning for that word before reading 'brachet'. A hart, by the way, is a kind of deer.

[20] 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. 421-422, 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 precision (which is easier to measure) suffices. Another is that *readability* measures are not altered by allowing "synonyms", but *comprehension* might be better demonstrated by allowing them. For further discussion of the limitations of cloze, see Kibby 1980.

reader's mind as are present in his [*sic*] own mind. Either he has some sense of what his readers know or he assumes that their knowledge stores resemble his" (Simon 1996: 171).

3. '*World* knowledge' connotes general knowledge about things *other* than the text's topic.

4. '*Domain* knowledge is specialized, subject-specific knowledge about the text's topic.

5. '*Commonsense* knowledge' connotes the culturally-situated beliefs that "everyone" has (e.g., that water is wet, that dogs are animals, maybe that Columbus discovered America in 1492, etc., but no "domain" knowledge). We include under this rubric both the sort of very basic commonsense information that the CYC[21] knowledge-representation and reasoning system is concerned with (Lenat 1995) 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.

Another aspect of this spectrum of prior knowledge is the expectations when encountering unknown words in reading situations such as a reading-comprehension passage on an SAT exam, reading a comic book, reading Shakespeare, or reading a recipe. Each of these affect prior knowledge and eliminate possibilities: In the SAT case, the reader expects (or should expect!) technical or obscure meanings. In the comic-book case, if 'kryptonite' is the unknown word, then the reader should know that it is more likely to be a nonce word than an obscure geological one. In the Shakespeare case, the reader should know that no unfamiliar word will refer to any modern contrivance such as email, but more likely to something in the context of the Elizabethan era. In the recipe case, unknown words often refer to food ingredients ('cumin') or cooking methods ('braising').[22]

3.4   The Proper Definition of 'Context'

Here is our first attempt to define the 'context' of *X*, with some caveats to be discussed in a moment:

Definition 1 The context of *an unfamiliar term X* **for** *a reader R is$_{def}$ the co-text of X + R's prior knowledge.*

Both co-text *and* prior knowledge are needed: 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; our emphasis)

most subjects infer that brachets are (probably) animals.[23]   But they do not make the inference solely from this 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: 255, our emphasis; cited by Kintsch 1980: 595). 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.)

---

[21]"CYC" is the name of 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/].

[22]Goldfain, personal communication.

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

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": They are "bridges" between the text and the reader's prior knowledge. And they do need to be inferred, though the inference involved is not (necessarily) deductive; rather, it is "abductive". Abduction is inference to the best explanation. [24] (It is non-deductive, because it is based on circumstantial evidence; thus, its conclusion can be false.) Thus, a reader might read in the text that a brachet bit a hart, abductively infer from prior knowledge 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.

We have preliminarily defined 'context' as a "sum" of information *from the text* and information *in the reader's mind*. The "text" (and hence "co-text") is something "out there" in the world; "prior knowledge" is something "inside" our heads, in our minds. But many cognitive scientists and reading specialists hold that, when you read, you "internalize" the text you are reading, i.e., you "bring it into" your mind (cf. G ärdenfors 1997, 1999ab; Jackendoff 2002 (§10.4), 2006; Rapaport 2003a).

Moreover, *this "internalized" text is more important than the actual words on paper*. As a simple example, consider the following dialogue in a *Betty* comic strip (16 April 2004):

> **Betty** (to her husband, who is reading): "Is that one of those 'I Spy' books?"
> **Husband:** "Yes, I'm using it to sharpen up my powers of perception."
> **Betty:** "Is it working?"
> **Husband:** "I've just started. I'm looking for five red pens."
> **Betty** (bending over to look at her husband's book): "That's 'hens.' Five red hens."
> **Husband:** "I perceive I'm not off to a very good start."

Here is a real-life example: One of us (WJR) read the sign on a truck parked outside one of our university cafeterias, 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? So he re-read the truck's sign more carefully. It actually said, "Mills *Welding* and Specialty *Gases*"! What matters for your understanding of the text is not what the text *actually* is, but what you *think* it is.

We need a name for this "internalized text". It is a "represented text", but 'representation' is one of those polysemous words itself in need of explication. It is also the reader's "mental model" of the text, but 'mental model' is a brand name (Johnson-Laird 1983) and best avoided. For now, we can't think of a better name than ... 'internalized text'.

So, our second approximation to a definition for the "context" of *X*—which resolves the "mind-body" duality of definition 1—is this:

Definition 2 **The context of** *an unfamiliar term X* **for** *a reader R is$_{def}$ R's* <u>internalized</u> *co-text of X + R's prior knowledge.*

But there's another problem: The internalized text "+" the prior knowledge might not be a simple "sum" (or "conjunction", or "union") of the two things. 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.

---

[24]The general form of an abduction is the deductive fallacy of "affirming the consequent": 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. §3.6, Hobbs et al. 1993.

the reader will automatically infer that 'he' refers to John (some say that 'he' and 'John' both refer to the same person; others say 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, e.g., 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 person *A* is shorter than person *B*, who is shorter than person *C*, then *A* is shorter than *C*; or 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 (again, see Ehrlich 1995, Rapaport & Kibby 2007).

In these cases, the whole is *greater* than the sum of the parts: The integration of the prior knowledge with the internalized text might include some extra beliefs that are not in the text and that were not previously in the 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*. (It's less likely, though possible—e.g., in the case of a typographical error—that you'd conclude that a sentence *in the text* was in error; cf. Rapaport 1991; Rapaport & Shapiro 1995, 1999.) In that case, you'll be revising your beliefs by *eliminating* something. Both of these forms of integration are components of reading comprehension.

So, that plus-sign in definitions 1 and 2 should be taken with a grain of salt. There is a whole subfield of AI, knowledge representation, and philosophy that studies this, called 'belief revision'. (See, e.g., Alchourrón et al. 1985, Martins & Shapiro 1988, Martins 1991, Gärdenfors 1992, Hansson 1999, Johnson 2006.) Here's a sample of some of their terminology applied to reading (but please also take some of this with a grain of salt, because the terminology isn't universally agreed on):

The plus-sign represents an operation that takes as input the reader's prior knowledge and internalized (co-)text, and that outputs an updated mental knowledge base that is a 'belief-revised integration' of the inputs. As the reader reads the text, some passages from it will be "*added*" (i.e., unioned or conjoined) to the reader's prior knowledge, and perhaps new inferences will be drawn; this is called 'expansion' of the prior knowledge base. Other text passages will be added, followed by the *elimination* of prior beliefs that are inconsistent with it (it is limited to prior beliefs, because a reader typically assumes that the text is correct, as just noted); this is called 'revision'. A few text passages (e.g., those involving typos) 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. (This, too, often happens when reading; cf. Rapaport 1991, Rapaport & Shapiro 1995.) After the text has been fully read, 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). Let's call the end result the '(belief-revised) integration' of the two inputs.

Pictorially, it might look like Figure 1. The left-hand rectangle represents the computational system's knowledge base or else the reader's mind; initially, it consists of (say) four propositions representing the reader's prior knowledge: PK1, PK2, PK3, and 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
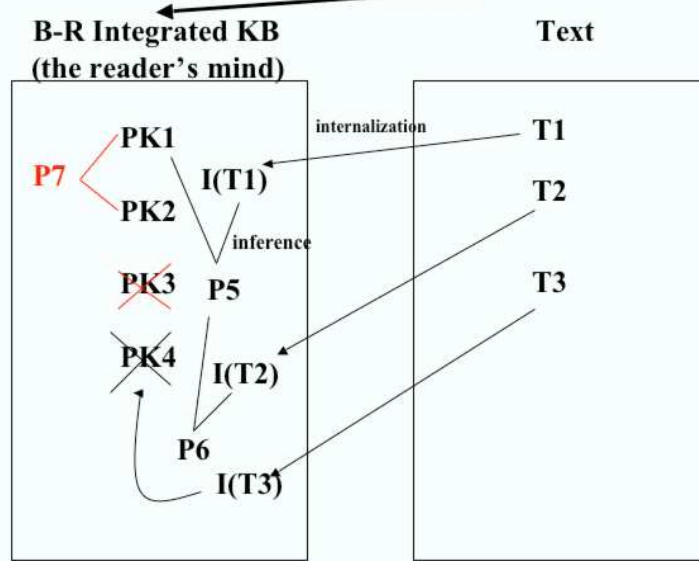
Figure 1: A belief-revised, integrated knowledge base and a text.

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).

The important point is that any "contextual" reasoning is done in the "context" *on the left-hand side*, i.e., in the belief-revised, integrated knowledge base, i.e., *in the reader's mind*. The context for CVA does not consist solely of the text being read (better: the co-text of the unfamiliar word) or of that (co-)text *together 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.

There's one final detail before we can present the formal definition: '*X*' was the unfamiliar term *in the text*. But we need a mental counterpart for it—an "internalized *X*"—because everything else has been internalized. So, our final definition of 'context' for CVA makes it a three-place relation among a reader, a term, and a text:

Definition 3
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 $T =_{def}$ *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) 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

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 is limited in the sense that it 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. (Diakidoy 1993: 3.)

> 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: 84–85; our emphasis.)

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. Our interpretation of the prevailing view of current CVA instructional materials and methods is that they privilege text as the source of all clues to an unknown word's meaning.

This perspective overlooks at least three other—and more significant and useful—CVA processes: text comprehension, prior knowledge brought to text comprehension by the reader, and reasoning processes. The computational linguist Jerry Hobbs (1990) argues that a text's meaning is a function of both the text *and* the reader's mind.[25] Hence, a meaning for a word is not usually given by the text alone.

Another limitation of current views of CVA is the notion that if CVA does not result at the moment of application in 'the correct' meaning of an unknown word, then CVA has failed. This view defies not only common sense about incremental learning, but also presumes that the purpose of CVA is solely vocabulary learning. In contrast, we argue that CVA is probably more useful to facilitate reading comprehension.

Our definition of 'context' also 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: The reader's mind is modeled by a knowledge base of 'prior knowledge' (including commonsense knowledge, world knowledge, perhaps some domain knowledge, etc.) expressed in a knowledge-representation language.

For us, that language is a semantic-network language (SNePS). As our computational cognitive agent (cf. Shapiro & Rapaport 1987, 1995; Shapiro 1989) reads the text, she (we have named 'her' 'Cassie') 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).

---

[25]This phenomenon seems akin to a similar one in the area of (visual) perception: Whether an object is seen as white is a function not only of the object in its environment but also of the perceiver's state of adaptation. If the perceiver has been in a red environment and has, accordingly, visually adapted to red, then an object that looks white to that perceiver will look red to someone entering the red environment from a white environment (Webster 2004).

As an example, consider the following series of passages (from Malory 1470: 66, 72) containing the unfamiliar word 'brachet':

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. (p. 66; my boldface, here and below.)

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.

In the presence of prior knowledge to the effect that:

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, and

PK6 only hounds bay,

Cassie outputs the following definition frame after processing ('reading') these sentences: [26]

```
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, and Possible Properties). The first slot has two fillers: a basic-level category (dog) and a superordinate-level category (hound). The second slot lists three actions that the only brachet that 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 'possible' properties: being valuable, being small, and being white; these properties are 'possible' in the same sense that the 'possible' actions are.

Thus, from our computational point of view, the 'context' that 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 that she integrated into it from the text. This matches our definition of 'context' for CVA. Cassie's definition is thus 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; cf. Rapaport 2002, 2003a).

(It should probably be mentioned at this point that 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 do. But there is no reason in principle that

---

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

she couldn't; this remains an open area of investigation. Other differences are that Cassie does not have to learn how to identify the printed form of words or match them to spoken forms, etc. We do not believe that these differences are significant for our purposes. Other researchers have investigated them computationally, however, so they are not outside the realm of possibility; see, e.g., Srihari et al. 2008.)

3.6 Distinguishing Internalized Co-Text and Prior Knowledge[27]

Although all relevant information is in this integrated knowledge base (i.e., in the reader's mind), there is sometimes a need to distinguish between beliefs that came from the (co-)text, beliefs that were already in the reader's prior knowledge, and beliefs that arose from inferences from both of these.

One simple case arises from the need to eliminate one of two inconsistent beliefs. To do this, we need to know their sources, so that we would know whether to retract a prior belief or a belief originating from the text. In our computational implementation, we do this by marking each proposition with a "knowledge category": 'story', meaning that the proposition came from the text; 'life', meaning that it came from prior knowledge, etc. (Ehrlich 1995, Martins & Cravo 1991, Rapaport & Ehrlich 2000). A human reader presumably has the text at hand, and can consult it to see if a proposition is explicitly stated in it; this is equivalent to Cassie's "story" label.

Another, more complex, case will also serve to illustrate the kind of prior knowledge that is useful for CVA. Consider the following text containing the (presumably) unfamiliar word 'detritus':

(T3) The birds alert nearby anglers that a massive school of menhaden is under attack by bluefish. The razor-toothed blues tear at the menhaden like piranhas in a killing frenzy, gorging themselves, some killing even when they are too full to eat, some vomiting so they can eat again. Beneath the blues, weak fish begin to circle, snaring the **detritus** of the carnage. (Franklin 2001.)

What prior knowledge might be useful for computing a meaning for 'detritus' from this passage? One possibility is the following "defeasible" rule (a defeasible statement is, roughly, one that can be "defeated"— i.e., rejected—later, on the basis of new information; i.e., it is contingent and can be contradicted or modified by later information):

(R) If fish x attacks fish y, and if fish z is weaker than fish x, then fish z will only get leftovers.

Where does this rule come from? Why 'leftovers', rather than something else? Because 'leftovers' is what at least one human reader of this passage inferred (in a verbal protocol) as the meaning of 'detritus', and we are trying to simulate that reader's understanding of the text. Thus, we assume that rule R must have been part of that reader's prior knowledge.

From rule R and the following part of text T3:

(T3.1) [W]eak fish begin to circle, snaring the detritus of the carnage.

we can reason that 'detritus' might be 'leftovers'. One way to do this is discussed below. But before showing *how* that inference can be made, we need a brief digression to explain why we *need* to show how.

This inference may seem vastly more complicated than is necessary. After all, either it is obvious to you why "detritus is leftovers" follows from text T3 and prior knowledge R, or it isn't. If it is, then you don't need the explanation. If it isn't, and if you are not a logician, the explanation to follow is likely to be unclear in the extreme. But this *is* the kind of reasoning done by any reader who infers "detritus is leftovers", even if it is done quickly, "incidentally", and unconsciously. A computer, however, would have to do it "deliberately" (or "consciously"). And by spelling out in detail how the inference is drawn, we can explain it to the reader. So, take a deep breath, and please proceed (or skip to §4!).

---

[27]This section is not essential to the central argument and can be omitted on first reading.

Consider the statements below, where (a) both formula R′—representing the formal version of rule R as expressed in the knowledge-representation language—and items PK1–PK5 come from the reader's prior knowledge, (b) 'weak-fish1' refers[28] to some item in the reader's prior knowledge that satisfies the conditions in PK3, and (c) T3.1′ is the formal version of T3.1):[29]

R′. $(\forall x, y, z)[(\text{Fish}(x)\ \&\ \text{Fish}(y)\ \&\ \text{Fish}(z)\ \&\ \text{Attacks}(x,y)\ \&\ \text{Weaker-than}(z,x))$
$\rightarrow\ \exists w[\text{Leftovers}(w)\ \&\ \text{Gets}(z,w)\ \&\ \forall v[\text{Gets}(z,v)\ \rightarrow\ v = w]]]$

This says: If there are three fish (call them x, y, and z), and if x attacks y, and if z is weaker than x, then there will be some leftovers (call them w) that z gets, and z doesn't get anything else (i.e., whatever z gets are those leftovers).[30]

PK1. Fish(bluefish)   (i.e., bluefish are fish)

PK2. Fish(menhaden)   (i.e., menhaden are fish)

PK3. Fish(weak-fish1) & Weaker-than(weak-fish1, bluefish)
(i.e., something—call it weak-fish1—is a fish and is weaker than bluefish)

PK4. $(\forall x, y)[\text{Tears-at}(x,y) \rightarrow \text{Attacks}(x,y)]$
(i.e., if you tear at something, then you attack it)

PK5. $(\forall x, y)[\text{Snares}(x,y) \rightarrow \text{Gets}(x,y)]$
(i.e., if you snare something, then you get it)

(PK4 and PK5 represent part of the reader's prior knowledge about the meaning of 'tear' and 'snare'.)

T3.1′. Begin-to-Circle(weak-fish1) & Snares(weak-fish1, detritus) & Tears-at(bluefish, menhaden)
(i.e., weak-fish1 begins to circle and snares the detritus, and the bluefish tear at the menhaden)

If we let x be bluefish, y be menhaden, and z be weak-fish1, then we can infer[31]

$$\exists w[\text{Leftovers}(w)\ \&\ \text{Gets}(\text{weak-fish1}, w)\ \&\ \forall v[\text{Gets}(\text{weak-fish1}, v)\ \rightarrow\ v = w]]$$

I.e., we can infer that there are leftovers that weak-fish1 gets and that the only thing that weak-fish1 gets are those leftovers. Similarly,[32] we can infer: Gets(weak-fish1, detritus) (i.e., weak-fish1 gets the detritus). Now, *if it were the case* that Leftovers(detritus) & $\forall v[\text{Gets}(\text{weak-fish1}, v) \rightarrow v = \text{detritus}]$ (i.e., the detritus is the leftovers, and there is nothing else for the weak fish to get except the detritus of the carnage), then we would be able to *deductively* infer the consequent of R′. So, we can *abductively* infer Leftovers(detritus) (see §3.5). This gives us a "meaning hypothesis" (or partial definition) for 'detritus': detritus is leftovers. This hypothesis is defeasible (i.e., it might be incorrect), yet plausible, and can serve as a first approximation

---

[28]Technically, 'weak-fish1' is a 'Skolem constant': It can be thought of as a temporary name, given to anything that will satisfy a certain description, when you don't know the real name of a thing that satisfies it. E.g., if I want to tell my students that anyone in the class who doesn't study will fail, I might say something like: 'Suppose someone in this class doesn't study; call him 'Fred'. Then Fred will fail." Here, 'Fred' is a Skolem constant. I also have to be sure that no one in the class is really named 'Fred'!

[29]Ignoring difficulties in representing generics like 'bluefish', 'menhaden', etc.

[30]For readers unfamiliar with the notation of predicate logic, '$\forall$' means 'for all', '&' means 'and', '$\rightarrow$' means 'if-then" (so '$P \rightarrow Q$' means 'if P, then Q'), '$\exists$' means 'there exists", constructions of the form '$P(t)$' are to be read as: 't is (or: is a) P", and constructions of the form '$P(t_1, t_2)$' are to be read as 't_1 stands in the P relation to t_2".

[31]I.e., by making those substitutions of constants for variables x, y, and z, we can apply *modus ponens* (see §6.2) to R′, PK1–PK4, and T3.1′ to get the conclusion.

[32]Instantiating and applying *modus ponens* to PK5 and T3.1′.

to a full definition. At the very least—but importantly—*it enables the reader to understand this passage* (at least to the extent that the reader understands what leftovers are).

Returning to our point about the need to distinguish prior knowledge from textual information, we *don't* want to infer from T3.1, *which is from the text,* and (e.g.) "They (those weak fish) also snared worms.", which let's suppose is *also* in the text, that 'detritus' are worms. One way to block this is to allow the previous inference to go through only when we use prior knowledge together with internalized text information, rather than two pieces of information from the text. And one way to do *that* is to associate each proposition with its source: text or prior knowledge (or an inference from these). As it happens, we already do this for independent reasons having to do with belief revision (as noted at the beginning of this section).

4   How to Do Things with Words in Context[33]

The story so far: (1) When we speak of figuring out a meaning for a word 'from context', we should mean: from the belief-revised integration of (a) the reader's prior knowledge with (b) the reader's internalized co-text of the word, where each proposition in this single, mental knowledge-base is marked with its source, and where we assume that the reader already understands all other words in the text. (2) There is a computer algorithm for doing this. And (3) we are adapting our algorithm to develop a better pedagogical curriculum than the current state of the art for *teaching* CVA.

Two often-cited papers by reading scientists (Beck et al. 1983; Schatz & Baldwin 1986) have claimed, not only that certain (textual) contexts are less than useful for doing CVA, but that most "natural" textual contexts (as opposed to artificial, "pedagogical" textual contexts) are not helpful at all. However, a careful examination of their arguments brings out several assumptions that are inconsistent with our computational theory of CVA. Thus, their objections (discussed below) do not apply to our theory. In short, it is possible to do lots of things with words in *any* (textual) context.[34]

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).[35]   We argue, by contrast, that *every* (textual) context contains *some* clues for constructing a meaning hypothesis. In this section, we examine their paper carefully to discern their assumptions and to see where our approach agrees and disagrees with theirs. (In what follows, except when quoting, 'textual context' refers to the co-text surrounding an unfamiliar word and 'wide context' refers 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 that enables the reader to *recognize* it as a clue ('clue' is a relative term).[36] Therefore:

Implication A1
    Clues in the text must be supplemented with other information in order to figure out a meaning.

---

[33] With apologies to Austin 1962.

[34] Alternatively, their claim *might* hold *at most* for learning meanings from purposefully-designed ("pedagogical") contexts, but not for CVA; cf. §2.3.3, above.

[35] All page references in this section are to Beck et al. 1983, unless otherwise noted.

[36] We owe this observation to our colleague Jean-Pierre Koenig.

This supplemental information must be supplied by the reader's prior knowledge. 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 (§2.3.2, above): Where he says "Guess", he must in fact mean "make an educated guess"—i.e., an inference—but that inference 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: (1) Not all of the reader's prior knowledge may be consciously available at the time of reading. (2) More importantly, each reader will bring to bear upon his or her interpretation of the text *idiosyncratic* prior knowledge (Dulin 1969, Garnham & Oakhill 1990, Rapaport 2003a).

(3) 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.)

(4) 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 a '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.[37]

(5) Variation also arises from *mis*interpretation (cf. Garnham & Oakhill 1990: 387ff), even simple misreading (see §3.4, above).

(6) Another source of variation, related to misreading, 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)".

4.1.2   Do Words Have Unique Meanings?

Beck et al.'s phrase 'the word's meaning' (177) reflects an assumption that is inconsistent with our theory:

---

[37]Shakthi Poornima, personal communication.

Assumption A2 A word has a *unique* meaning.

The definite description 'the word's meaning' or 'the meaning of a word' suggests incorrectly that a word has a unique meaning (see §2.1). To be charitable, we could say that what's normally intended by this phrase is 'the meaning of a word *in the present context*" (recall Deighton's observation that context determines meaning; see §3.1, above). But it follows from our observations about implication A1 (i.e., that textual clues need to be supplemented with other information) 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, unwarranted assumption is:

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

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).

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, and it might very well be the case that the *author's* intended meaning is *not* thus determined. Our investigations suggest that this is almost always the case. 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* 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 other readers or from the author's intended meaning—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 [sic] 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 the reader can figure out *a* meaning for the word *that is sufficient to enable him or her to understand the text well enough to continue reading*.

Clarke & Nation (1980: 213) 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. This suggests a difference in attitudes towards CVA of L1 reading educators such as Beck et al., who expect "precise" and "correct" meanings, and L2 educators such as Nation, who don't.) 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"[38] 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 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.[39]

Is an impasse a disagreement? One dictionary defines it as a "deadlock" (Waite 1998). Suppose that "deadlock" is "the correct meaning". If the reader decides that 'impasse' means "disagreement", not "deadlock", has the reader misunderstood the passage? Consider the following scenarios:

1. The reader never sees the word 'impasse' again. It then hardly matters whether she has "correctly" understood the word (though she has surely figured out *a* very plausible meaning).

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

3. The reader sees the word again, in a context in which "disagreement" is *not* a plausible meaning, but "deadlock" is. E.g., she might read a computer science text discussing operating-system deadlocks, in which a particular deadlock is referred to as an "impasse". Here, 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:

---

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

[39]Plausibly, if agreement is *im*possible, then *dis*agreement *is* possible. And, plausibly, 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 (recall §3.6), but they are the sort of inferences our protocols show that readers actually make.

i. She decides that she must have been wrong about 'impasse' meaning "disagreement", and she 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 §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 (§2.3.3), 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 two kinds: *pedagogical* and *natural*. The former are "specifically designed for teaching designated unknown words" (178). It will be of interest later 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.[40] 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:

---

[40]Many authors write of "cues"; others, of "clues"; some (e.g., Beck et al.), 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 to (perhaps consciously) infer something. Thus, not all cues are clues, and not all clues are cues. The two terms seem to be interchangeable in the literature, but we will try to use them in the way mentioned here, except when quoting.

> 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. But this is a minor matter; surely, many such allegedly misdirective co-texts could be found in nature, so to speak.

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

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

But, in the present example, there might be other clues, preceding or following this short co-text, that would rule out "admiringly". Perhaps we know or could infer from earlier or later passages that Ginny is jealous of Sandra, or that she is inclined to ironic comments. Strictly speaking, one could logically infer from this passage 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) But, 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, primarily 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 passages that are relevant for CVA? We take a holistic view of meaning; thus, all passages are potentially relevant (Rapaport 2002). The issue that our definition algorithms help resolve is how to filter out a dictionary-like definition from this wealth of data; cf. 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 only occurs 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 the reader concludes or whether the reader concludes anything at all, for it is highly unlikely that anything crucial will turn on such a word. 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 for *learning* a meaning for a word in a pedagogical situation (see §2.3.3, 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. Rather, we are developing a method to assist readers in hypothesizing meanings in a way that facilitates subsequent 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 it means' (Kibby et al., forthcoming; 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, neither of these further assumptions is necessarily the case. In real life, these are unavoidable problems. However, we expect that 'word consciousness' grows with frequent 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 encounters.

4.1.5.2 Nondirective Co-Texts. The next category is 'nondirective contexts, which 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.)

Again, the evidence is underwhelming, and for the same reasons: 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 context; it is not that their contexts are mis-directive or non-directive that makes them so (see §6.6, below).

The reader can ignore a single unfamiliar word in *both* misdirective *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 §4.1.1). As a literary conceit, it might be excusable; in expository writing, it would not be.

4.1.5.3 Syntactic Manipulation. But even misdirective and non-directive co-texts are capable of yielding a clue. The technique for squeezing a clue out of any co-text is to syntactically manipulate 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 (see §7.6).

From the above 'misdirective' text, we could infer that, whatever else 'grudgingly' might mean, it could be defined (albeit only vaguely) as 'a way of saying something' (and we could list all sorts of such ways, and hypothesize that 'grudgingly' is one of them). Moreover, it could be defined (still vaguely) as 'a way of (apparently) praising someone's performance' (and we could list all sorts of 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.[41] 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.

---

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

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.)

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:[42]

> Unlike his brothers, who were noisy, outgoing, and very talkative, Fred was quite *taciturn*. (Dulin 1970.)[43]

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 'tatterdamlion' in §2.1.)

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). But, here, their example is that of 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.[44] But, more importantly, the fact that it is a noun suggests that it is not so much the co-text that is helpful as it is the fact that it 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-Like Tasks

Beck et al. conducted an experiment involving subjects who were given 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). Independent of the results, there are several problems with this set-up:

(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.)

(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:

---

[42] 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; contextually analyzed in Anger 2003).

[43] 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).

[44] The content, spelling of 'leapt', and occurrence of 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", online at
[http://www.nifl.gov/partnershipforreading/publications/reading_first1vocab.html].

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

Our project, by contrast, is focused on deliberate CVA, carefully taught and practiced.
   (4) Another problem arises from the next unwarranted assumption:

Assumption A8 Cloze-like tasks are a form of CVA.

Schatz & Baldwin (1986: 450) also claim that "Using context to guess the meaning of a semantically unfamiliar word is essentially the same as supplying the correct meaning in a cloze task." But this is not the case: In cloze-like tasks, a word in a passage is replaced with a blank, and the reader is invited to guess (rather than figure out) the missing word. But this is not CVA, and there is a unique, correct answer. (Moreover, cloze is not valid as a measure of reading comprehension; Kibby 1980.) In CVA, the goal is to figure out a meaning that is sufficient for understanding the passage.

   Here, a serious methodological difficulty faces all CVA researchers: If you want 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?), but when obvious neologisms or blank spaces are used, readers focus on "getting" the word, not on expressing its possible meaning. These tasks are probably related, but they are distinctly different, too.

   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. (Wolfe 2003 discusses the related problem of inventing names in social science.) 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 (see Gardner 2007: 337, 344, 347, for related observations), nor do we have any clever solutions. Our preferred technique for now is to use a plausible-sounding neologism (with appropriate affixes) and then to inform the subject that it is a word from another language that might or might not have a single-word counterpart in English, but that in any case 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, however, "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). [45]

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] … , 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* an efficient mechanism 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 that CVA is not 'an effective strategy for inferring word meanings" (440). In contrast, we argue that *wide* context *can* help you figure out *a* meaning for an unfamiliar word, so that CVA *is* an effective strategy for inferring (better: figuring out) word meanings.

As with Beck et al., 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. 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). More importantly, as Schwartz (1988: 111) points out, *CVA needs to be applied to the task of understanding a dictionary definition*

---

[45]All page references in this section are to Schatz & Baldwin 1986, unless otherwise noted.

itself, which is, after all, merely one more co-text containing the unfamiliar word (cf. Gardner 2007: 342). CVA is the base case of a recursion one of whose recursive clauses is "look it up in a dictionary".

Guessing, rather than inferring, might be close to instantaneous. We have argued that CVA must be more than a process of (merely) guessing. Anyone can guess, but, if the reader is expected to search carefully through the surrounding text for clues to a word's meaning, then there can—and must—be specific tasks that can be undertaken to make use of those clues for constructing a possible meaning (see §7.7). But Schatz & Baldwin (and many others) are inconsistent in their terminology. For instance, we read on p. 441, following a discussion of easily guessable cloze passages, that "These studies do show that readers can *infer* the meanings of words." But there is no inference (other than immediate, unconscious inference) involved in figuring out that the blank in "He drank a ___ of coffee" is probably a quantity-word, e.g., 'cup', 'mug', or possibly 'gallon'. A guess really is all that's needed here.

4.2.2 Schatz & Baldwin's Methodology

Schatz & Baldwin offer the results of several experiments to support their claims. There are a number of problems with their methodology (or, at least, with their *description* of their methodology).

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. Some of the words they chose were: 'cogently', 'cozened', 'ignominiously', 'imperious', 'inexorable', 'perambulating', 'recondite', 'salient' (442). Note that four (or 50%) are adjectives, two (25%) are adverbs, one (12.5%) is a verb ('cozened'), 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 75% of the unfamiliar words are modifiers, known to be among the most difficult of words to learn meanings for. Schatz & Baldwin's example passages consist of an adverb ('ruefully'), three adjectives ('glib', 'pragmatic', 'waning'), and four nouns ('yoke', 'coelum', 'dearth', 'ameliorating'). This brings the statistics to around 67% for adjectives and adverbs, 27% for nouns, and 6% for verbs (not counting 'perambulating'). 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) than they do about weaknesses of contexts. (In longer texts, there may be more opportunities to hypothesize meanings for modifiers.)

4.2.2.2 CVA vs. Word-Sense Disambiguation. Moreover, in two of Schatz & Baldwin's 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 there is of figuring out a meaning, for the simple reason that there will be a minimum of textual clues. The larger the co-text, the greater the chance, for the simple reason that a large enough co-text might actually include a definition of the word! (Recall from §4.2.1, however, 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 in which the unknown word occurs) 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. An 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.

Yet 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 is something that we all do; it is the best explanation for how we learn most of our vocabulary (see §2.1, 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? This is an open, empirical question. In principle, the earlier, the better.

4.2.3   Three Questions about CVA

In their general-discussion section, Schatz & Baldwin raise three questions (447; emphasis in original):

(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 to this question is: Both traditional (§2.3.4) and non-traditional (§6.6) context clues do occur and 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?" But "accuracy" is also irrelevant. Moreover, a "denotation" (in the sense of an external referent of a word) is best provided by demonstration or by a graphic illustration, and a "connotation" (in the sense of 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 to this question 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 (§4.1.5.3).

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 it is. 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 (even if only minimal) to a word's meaning.

More precisely: For every textual context C, and for every word w in C, w has a meaning in C (at the very least, its "algebraic" meaning obtained by rephrasing C to make w the subject). But w will also have a meaning that is partly determined by the reader R (and his or her accessible prior knowledge). And, because the reader's prior knowledge may be time-dependent, we need to add "at time t" to this formulation. Thus, for every w, C, R, t, if w is in C, and R reads C at t, then w has a meaning in C for R at t. Less formally, every time you read a word in a text, it will have a meaning for you that is determined by the text integrated with your prior knowledge at that time. And, of course, none of the meanings that w has for R is necessarily "the" meaning (in either a dictionary sense or that of a reading teacher).

CVA.2   The surrounding 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.

Corollary CVA.2.1 There is no such thing as a "good" or "bad" co-text *simpliciter* (or a "misdirective", "non-directive", "general", or "directive" co-text, either); the value of a co-text depends in part on the reader's prior knowledge and ability to use clues and prior knowledge together, and in part on the presence (or absence) of *potential* clues.[46]

CVA.3 CVA is distinct from cloze-like tasks and from 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.

CVA.5 Many contexts may be required for CVA to "asymptotically" approach 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 A word does not have a (single) correct meaning, not even in directive and pedagogical contexts.

Alternative 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).

Corollary 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.

(This often happens 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 delving into our computational theory, let's consider one published protocol of a (presumably secondary-school-aged) reader ("Marian") figuring out what an unknown word might mean (Harmon 1999: 328). 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)

---

[46]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*.

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 even two of them that she has conflated. Cassie (our computer system), however, has to do it "consciously", which means that we programmers have to explicitly tell her how to do it (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's the inference that Marian makes:

1. 'Davidson-Bart' is a proper name.

   (This is prior knowledge, not 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.

   (This comes 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 if there were 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.

   (This is an inference from prior knowledge (1) of the kind of entity that might need legal counsel and (2) that names are likely to be names 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 in the text. Actually, it's an inference from the actual text, which says "Davidson-Bart is a multinational conglomerate". But prior linguistic knowledge tells us that if something is a multinational conglomerate (whatever that means), then it's probably a conglomerate. This is defeasible (§3.6): 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:

37

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).

The word 'multinational' could play a couple of roles. First, it's 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). Second, 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 the case that our computational cognitive agent Cassie 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 initially supplied with a 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.) In SNePS, there are three basic ways to express prior knowledge: Call them (1) 'basic propositions', (2) 'proposition-based rules', and (3) 'path-based rules'.

Examples of *basic propositions* include 'Someone is named John', 'Someone is tall', 'Someone likes someone (else)', 'Some particular kind of thing belongs to someone', etc. (see §6.5, below). In general, 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 atomic sentences of the form Px or Rxy, for instance, 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'.

(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 (logicians call this rule of inference '*modus ponens*').

(3) *Path-based rules* generalize the inheritance feature of semantic networks, enabling Cassie to infer, e.g., that Fido is an animal, if she believes that Fido is a brachet, that brachets are dogs, and that dogs are animals, or, e.g., to believe that Fido has fur, if she believes that animals have fur and if she believes (or can infer) 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. (Cf. Shapiro 1991. This is all a vast oversimplification, but will suffice for now.)

6.3 Special Rules

Prior knowledge can be of a wide variety (§3.3). We try to limit the prior knowledge to propositions that are necessary for Cassie to understand the meaning of 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 meaning postulates about the crucial terms in the co-text, i.e., necessary *or* sufficient conditions concerning these terms), there is need for rules of a very special and general sort. We saw one example in §6.1; here are a few more examples:

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 agent y on object 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 (§3.6). 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 cited in Ehrlich 1995 incorporated all prior knowledge from previous experiments.) In general, Cassie would acquire her prior knowledge from reading, being told, previous reasoning, etc.; in short, she would learn it in any of the variety of ways that one learns anything (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 them, they seem unmotivated.

6.5 Format of Prior Knowledge

Armed with her prior knowledge, Cassie begins to read the text. We input the text to a computational grammar, which outputs a semantic representation of the text. Currently, the grammar is implemented in an augmented-transition-network formalism (Woods 1970, Shapiro 1982). The output consists of a semantic

network in the SNePS formalism. (When our full system is implemented, she 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. (Where these prove insufficient, we extend the set. 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. The assumption is that the meaning of a complex sentence should be the (combined) meanings of the shorter sentences into which it gets analyzed. In each of the entries 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 (for details, see Rapaport 2006b; for more information on SNePS, see Rapaport 2007).

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). Whether it was already in the network or newly assimilated, Cassie then does "forward" inference on it. This models Cassie as a reader who thinks about each sentence that she reads. 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). 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 noun or verb; call it X. If X is not in Cassie's lexicon (because she has never read the word, not even in the current text), she will respond with "I don't know". If X *is* in her lexicon, then—whether or not X occurs in the current text (though typically, of course, it will)—Cassie will search her entire network for a subset of the information that can fill in the slots of a definition frame. In the default case, Cassie will "algebraically/syntactically" manipulate the only sentence containing X so that X becomes its subject. In general, Cassie will look for general, basic-level information, though in its absence she will report information specific to known instances of X. Each of these steps is repeated for subsequent occurrences of X, until a stable definition is reached.

The noun and verb algorithms operate by searching through the belief-revised, integrated knowledge base (the "wide context") for information that can be incorporated into the 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 having 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. (For details, see Ehrlich 1995 and 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"; on the notion of basic-level categories, see, e.g., Rosch 1978, Mervis & Rosch 1981); if it fails to find or infer any, it seeks most-specific-level class memberships; if it fails to find or infer any of these, it seeks names of individuals (e.g., it might decide that it doesn't know what kind of thing a brachet is, but it might know that 'Fido' is the name of one)

- properties of Xs (size, color, etc.); if it can't find properties that it believes are exemplified by all Xs,

then it seeks properties of individual Xs—these are considered to be "possible properties" of Xs in the sense that our known X exemplifies it, so it is "possible" that some Xs exemplify them

- structural information about Xs (part-whole, physical structure, etc.); if it can't find structural information that it believes is exemplified by all Xs, then it seeks "possible" structural information exemplified by individual Xs

- acts (or "possible" acts) that Xs perform or that can be done to, or with, Xs

- agents that do things to, or with, Xs, or to whom things can be done with Xs, or that own Xs

- possible synonyms and antonyms of the word X.

I.e., 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.

In a similar manner, the verb algorithm deductively searches the knowledge base for:

- class membership information (e.g., based on Schank & Rieger's (1974) "Conceptual Dependency" classification scheme developed in AI, or Levin's (1993) classification scheme developed in cognitive linguistics): 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, slow walking)

- transitivity or subcategorization information (i.e., is there only an agent of the act, or are there also 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 harder; we can, however, produce definitions of certain adjectives. This is an area of current research (see §7). These are very much more difficult to figure out a meaning for unless there is very specific kinds of information in the context. After all, if the word '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 the word 'salty', though it could be the word '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.

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 (see, e.g., Quillian 1967, Rapaport 1981). Later modifications have been based on protocols of human readers doing CVA (reported in 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, on the other hand, get bored, are forgetful, and don't always draw every relevant logical consequence. For instance, 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 light bulb goes on, 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, weren't thinking about it, or failed to draw the inference until it was pointed out to them. Cassie always infers such things. The important point, however, is not so much 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.

Moreover, the simulation is implemented in a symbolic AI system, not a statistically-based or connectionist system (recall §1, above). This means that 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' in the sense that it would be thinking in some kind of rigid, uncreative, 'mechanical' way. 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 what we *can* do is devise a curriculum for teaching CVA that is a human adaptation of our rule-based algorithms. A statistically-based algorithm would not be able to 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. There are, however, some things that 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 a set of rules (an algorithm, or strategy, if you prefer) that a human reader can always rely on when at a loss for what to do. Finally, the human-oriented CVA strategies must be embedded in a curriculum that begins with examples and instructions provided by someone familiar with the technique, moves to assisted practice, and evolves into a tool that the reader can rely on in future reading.

We call the full curriculum 'Contextual Semantic Investigation', which not coincidentally has the currently popular abbreviation 'CSI' (Kibby et al., in press). This emphasizes one of the two guiding metaphors for our curriculum: 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).

The 'scaffolding' in which the curriculum is embedded begins with the teacher modeling CSI, 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 to 'solve the crime'. Next, the burden is placed on the students, who attempt CSI with the teacher's help, followed by small groups of students working together. Finally, each student is given an opportunity to work on his or her own.

7.1 The Basic Algorithm

The basic human-centered algorithm consists of two main steps:

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:

    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: Readers often skim right over an unfamiliar word, either not paying any attention to it at all, or else hoping that it won't turn out to be important. So the first step is to bring the word to the reader's awareness and to make the reader see the need for understanding it. This is easier to do in a classroom setting (where the teacher can simply tell the students that they need to figure out what the word means) than it is when one is reading on one's own. But by the end of classroom instruction, readers should have become more aware of unfamiliar words. (There is some empirical support that this does occur, e.g., Beck et al. 1982, Christ 2007.) Step 7.1.2a allows the reader to expand the textual context under consideration, if needed. The process ends when the reader has hypothesized a meaning that is consistent with both the text and his or her prior knowledge.

7.2 Testing the Hypothesis

Testing the hypothesis is straightforward, but must be done. Otherwise, a poor understanding of a word's meaning can lead to further misunderstanding later in the text. Testing is 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).

Note that if *X* is, say, a noun, but the reader has construed its definition in, say, verb form, then the substitution shouldn't make sense,[47] and the student will have to revise *H*. This kind of revision should be fairly straightforward with the teacher's help.

7.3 Generating a Hypothesis

The bulk of the work, of course, lies in *generating* a meaning hypothesis. Here, too, our curriculum differs from our computer algorithm by giving the reader a chance to make a guess. As we noted in §2.3.2, above, computers can't (easily) guess. 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, and 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.)

---

[47] 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".

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 Making an Educated Guess

Making an 'educated' guess is not merely guessing, but is the result of careful and active thinking.

To make an "educated" guess:

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

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

But what does the poor reader do, who has not yet succeeded in generating a hypothesis?

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 the 'algebraic' manipulation mentioned above.

To "solve for X":

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

E.g., the sentence "A hart ran into King Arthur's hall with a *brachet* (*X*) next to him" would be syntactically manipulated 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 Searching 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 Xing is
   - what kinds of acts are Xings
   - properties of Xing (e.g., manner)
   - transitivity
   - look for agents and objects of Xing
   - 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 Constructing 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:
   (a) What (kind of thing) is X?
   (b) What is it like?
   (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' shown in §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 generally considered to be "AI-complete" (Shapiro 1992). I.e., 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 there for those who need them.

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

Acknowledgments

References

Aist, G. (2000), "Helping Children Learn Vocabulary during Computer Assisted Oral Reading", *Proc. 17th Nat'l. Conf. Artif. Intell.* (Menlo Park, CA, & Cambridge, MA: AAAI/MIT Press): 1100–1101.

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", *Reading Res. Qtly.* 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* (Erlbaum): 243–257.

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, Thomas (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", *Amer. Ed. Res. J.* 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 Res. Qtly.* 37: 150–176.

Beals, D.E. (1997), "Sources of Support for Learning Words in Conversation", *J. Child Lang.* 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", *Elem. School J.* 83: 177–181.

Beck, I.L.; McKeown, M.G.; & Omanson, R.C. (1984), "The Fertility of Some Types of Vocabulary Instruction", paper presented at Amer. Educational Res. Assn. (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", *J. Ed. Psych.* 74: 506–521.

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

Bowey, J.A.; & Muller, D. (2005), "Phonological Recoding and Rapid Orthographic Learning in Third-Graders' Silent Reading", *J. Experimental Child Psych.* 92: 203–219.

Brown, G., & Yule, G. (1983), *Discourse Analysis* (Cambridge, UK: Cambridge U. Press).

Bruner, J. (1978), "Learning How to Do Things with Words", in J. Bruner & A. Garton (eds.), *Human Growth and Development* (Oxford: Oxford U. Press): 62–84.

Brysbaert, M.; Drieghe, D.; & Vitu, F. (2005), "Word Skipping", in G. Underwood (ed.), *Cognitive Processes in Eye Guidance* (Oxford: Oxford U. Press): 53–78.

Budiu, R., & Anderson, J.R. (2001), "Word Learning in Context: Metaphors and Neologisms", *Tech. Rep. CMU-CS-01-147* (Carnegie Mellon U., School of Comp. Sci.).

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

Buikema, J.L., & Graves, M.F. (1993), "Teaching Students to Use Context Cues to Infer Word Meanings", *J. 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].

Catford, J.C. (1965), *A Linguistic Theory of Translation* (Oxford: Oxford University Press).

Christ, T. (2007), "Oral Language Exposure and Incidental Vocabulary Acquisition", Ph.D. diss. (Buffalo: SUNY Buffalo Dept. of Learning & Instruction).

Clancey, W.J. (2006), "Scientific Antecedents of Situated Cognition", in P. Robbins & M. Aydede (eds.), *Cambridge Handbook of Situated Cognition* (New York: Cambridge, forthcoming).

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

Cunningham, A.E., & Stanovich, K.E. (1998), "What Reading Does for the Mind", *Amer. 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", *Comm. 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", Ph.D. diss. (Urbana-Champaign: U. 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", *Reading Teacher* 23: 440–445, 469.

Ehrlich, K. (1995), "Automatic Vocabulary Expansion through Narrative Context", *Tech. Rep. 95-09* (Buffalo: SUNY Buffalo Dept. Computer Science).

Ehrlich, K. (2004), "Default Reasoning Using Monotonic Logic", *Proc. 15th Midwest AI & Cog. Sci. Conf.*: 48–54.

Ehrlich, K., & Rapaport, W.J. (1997), "A Computational Theory of Vocabulary Expansion", *Proc. 19th Ann. Conf. Cog. Sci. Soc.* (Mahwah, NJ: Erlbaum): 205–210.

Ehrlich, K., & Rapaport, W.J. (2004), "A Cycle of Learning", *Proc. 26th Ann. Conf. Cog. Sci. Soc.* (Erlbaum, 2005): 1555.

Elman, J.L. (2007), "On Words and Dinosaur Bones", in D.S. McNamara & J.G. Trafton (eds.), *Proc. 29th Ann. Conf. Cog. Sci. Soc.* (Austin, TX: Cog. Sci. Soc.): 4.

Engelbart, S.M., & Theuerkauf, B. (1999), "Defining Context within Vocabulary Acquisition", *Lang. Teaching Res.* 3: 57–69.

Etzioni, O. (ed.) (2007), *Machine Reading* Tech. Rep. 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), "The Most Important Fish in the Sea", *Discover* 22(September): 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* (Oxford: Blackwell, 1970): 56–78.

Fukkink, R.G., & De Glopper, K. (1998), "Effects of Instruction in Deriving Word Meaning from Context", *Rev. Ed. Res.* 68: 450–458.

Gärdenfors, P. (ed.) (1992), *Belief Revision* (Cambridge, UK: Cambridge U. Press).

Gärdenfors, P. (1997), "Meanings as Conceptual Structures", in M. Carrier & P. Machamer (eds.), *Mindscapes* (U.

Pittsburgh Press).

Gärdenfors, P. (1999a) "Does Semantics Need Reality?", in A. Riegler, M. Peschl, & A. Von Stein (eds.), *Understanding Representation in the Cognitive Sciences* (Dordrecht, the Netherlands: 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 Psych.* 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", *Cogn. & Brain Theory* 4: 161–178.

Gentner, D. (1982), "Why Nouns Are Learned before Verbs", in S.A. Kuczaj (ed.), *Language Development* (Erlbaum) 2: 301–334.

Gildea, P.M.; Miller, G.A.; & Wurtenberg, C.L. (1990), "Contextual Enrichment by Videodisc", in D. Nix & R. Spiro (eds.), *Cognition, Education, and Multimedia* (Hillsdale, NJ: Erlbaum): 1–29.

Gillette, J.; Gleitman, H.; Gleitman, L.; & Lederer, A. (1999), "Human Simulations of Vocabulary Learning", *Cogn.* 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", *Proc. 5th Int'l. Joint Conf. AI* (Los Altos, CA: William Kaufmann): 67–68.

Guha, R.V. (1991), "Contexts", Ph.D. diss. (Stanford U.)

Haastrup, K. (1991), *Lexical Inferencing Procedures or Talking about Words* (Tübingen: Gunter Narr).

Halliday, M.A.K. (1978), *Language as Social Semiotic* (Baltimore: University Park Press).

Hansson, S.O. (1999), *A Textbook of Belief Dynamics* (Dordrecht, the Netherlands: Kluwer).

Harmon, J.M. (1999), "Initial Encounters with Unfamiliar Words in Independent Reading", *Res. in the Teaching of Eng.* 33: 304–318.

Harris, A.J., & Sipay, E.R. (1990), *How to Increase Reading Ability, 9th edition* (White Plains, NY: Longman).

Harsanyi, D. (2003), "The Old Order", *Nat'l. Rev.* (May 5): 45–46.

Hastings, P.M., & Lytinen, S.L. (1994a), "The Ups and Downs of Lexical Acquisition", *Proc. 12th Nat'l. Conf. AI* (Menlo Park, CA: AAAI/MIT Press): 754–759.

Hastings, P.M., & Lytinen, S.L. (1994b), "Objects, Actions, Nouns, and Verbs", *Proc. 16th Ann. Conf. Cog. Sci. Soc.* (Hillsdale, NJ: Erlbaum): 397–402.

Haugeland, J. (1985), *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT).

Higginbotham, J. (1985), "On Semantics", in E. Lepore (ed.), *New Directions in Semantics* (Academic, 1987): 1–54.

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", *Amer. Educator* 27(1): 10–13, 16–22, 28–29, 48.

Hirst, Graeme (2000), "Context as a Spurious Concept", *Proc., Conf. Intell. Text Processing and Comp. Ling. (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. Intel.* 63: 69–142.

Horst, S. (2003), "The Computational Theory of Mind", in E.N. Zalta (ed.), *Stanford Encyc. Phil.* [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* (Blackwell): 349–381.

Ide, N.M., & Veronis, J. (eds.) (1998), Special Issue on Word Sense Disambiguation, *Comp. Ling.* 24(1).

Iwańska, Ł. & Zadrozny, W. (1997), Special Issue on Context in Natural Language Processing, *Comp. Intel.* 13(3).

Jackendoff, R. (2002), *Foundations of Language* (Oxford: Oxford U. Press).

Jackendoff, R. (2006), "Locating Meaning in the Mind (Where It Belongs)", in R.J. Stainton (ed.), *Contemporary Debates in Cognitive Science* (Malden, MA: Blackwell): 219–236.

Johnson, F.L. (2006), "Dependency-Directed Reconsideration", Ph.D. diss. (Buffalo: SUNY Buffalo Dept. Comp. Sci.

    & Eng'g.) [http://www.cse.buffalo.edu/sneps/Bibliography/joh06a.pdf]

Johnson-Laird, P.N. (1983), *Mental Models* (Cambridge, UK: Cambridge U. Press).

Johnson-Laird, P.N. (1987), "The Mental Representation of the Meanings of Words", *Cogn.* 25: 189–211.

Jurafsky, D.; & Martin, J.H. (forthcoming), *Speech and Language Processing, 2nd Ed.* (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", *J. Adolescent and Adult Literacy* 39: 208–223.

Kibby, M.W. (2007), "Lean Back and Think", *.edu Alumni Newsletter of the UB Grad. Sch. Ed.* (Spring): 13,20 [http://www.gse.buffalo.edu/newsletter/GSE_newsltr_sp07.pdf].

Kibby, M.W.; Rapaport, W.J.; & Wieland, K.M. (2004), "Helping Students Apply Reasoning to Gaining Word Meanings from Context", *49th Ann. Conv. Int'l. Reading Assn.*.

Kibby, M.W.; Rapaport, W.J.; Wieland, K.W.; & Dechert D.A. (forthcoming), "CSI: Contextual Semantic Investigation for Word Meaning", in L.A. Baines (ed.), *Literacy through Multisensory Learning* (Alexandria, VA: Assn. Supervision and Curriculum Development).

Kintsch, W. (1980), "Semantic Memory", in R.S. Nickerson (ed.), *Attention and Performance VIII* (Erlbaum): 595–620.

Kintsch, W. (1988), "The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model", *Psych. Rev.* 95(2): 163–182.

Kintsch, W. & van Dijk, T.A. (1978), "Toward a Model of Text Comprehension and Production", *Psych. Rev.* 85: 363–394.

Kuhn, M.R., & Stahl, S.A. (1998), "Teaching Children to Learn Word Meanings from Context", *J. Lit. Res.* 30: 119–138.

Lakoff, G. (1987), *Women, Fire, and Dangerous Things* (Chicago: U. Chicago Press).

Lammert, A. (2002), "Defining Adjectives through Contextual Vocabulary Acquisition" [http://adamlammert.com/school/cva/cva_report.pdf].

Langacker, R. (1999), *Foundations of Cognitive Grammar* (Stanford, CA: Stanford U. Press).

Lenat, D.B. (1995), "CYC: A Large-Scale Investment in Knowledge Infrastructure", *Comm. ACM* 38(11): 33–38.

Lenat, D.B. (1998), "The Dimensions of Context-Space" [http://www.cyc.com/].

Levin, B. (1993), *English Verb Classes and Alternations* (Chicago: U. Chicago Press).

Loui, M.C. (2000), "Teaching Students to Dream" [http://wocket.csl.uiuc.edu/~loui/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.P., & Cravo, M.R. (1991), "How to Change Your Mind", Noûs 25: 537–551.

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

McCarthy, J. (1993), "Notes on Formalizing Context", *Proc. 13th Int'l. Joint Conf. AI* (San Mateo, CA: Morgan Kaufmann): 555–560.

McKeown, M.G. (1985), "The Acquisition of Word Meaning from Context by Children of High and Low Ability", *Reading Res. Qtly.* 20: 482–496.

Mervis, C.B.; & Rosch, E. (1981), "Categorization of Natural Objects", *Ann. Rev. Psych.* 32: 89–115.

Miller, George A. (1985), "Dictionaries of the Mind", *Proc. 23rd Ann. Meeting Assn. Comp. Ling. (U. Chicago)* (Morristown, NJ: Assn. Comp. Ling.): 305–314.

Miller, G.A. (1986), "Dictionaries in the Mind", *Lang. and Cog. Procs.* 1: 171–185.

Miller, G.A. (1991), *The Science of Words* (Sci. Amer. Library).

Miller, G.A., & Gildea, P.M. (1985), "How to Misread a Dictionary", *AILA Bull.* (June): 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 Ling.* 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, Rev. Ed.* (Bantam).

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

Nagy, W.E., & Anderson, R.C. (1984), "How Many Words Are There in Printed School English?", *Reading Res. Qtly.* 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* (Hillsdale, NJ: Erlbaum): 19–35.

Nagy, W.E.; Herman, P.A.; & Anderson, R.C. (1985), "Learning Words from Context", *Reading Res. Qtly.* 20: 233–253.

Nation, I.S.P. (2001), *Learning Vocabulary in Another Language* (Cambridge U. Press).

Nation, P., & Coady, J. (1988), "Vocabulary and Reading", in R. Carter & M. McCarthy (eds.), *Vocabulary and Language Teaching* (London: Longman): 97–110.

Newell, A.; Shaw, J.C.; & Simon, H.A. (1958), "Elements of a Theory of Human Problem Solving", *Psych. Rev.* 65: 151–166.

Ong, W.J. (1975), "The Writer's Audience Is Always a Fiction", *PMLA* 90(1): 9–21.

Pelletier, F.J. (2001), "Did Frege Believe Frege's Principle?", *J. Logic, Lang. and Info.* 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* (New York: NYU): 148–179.

Quillian, M.R. (1967), "Word Concepts", *Behavioral Sci.* 12: 410–430.

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

Quine, W.V.O. (1961), "Two Dogmas of Empiricism", in W.V.O. Quine, *From a Logical Point of View, 2nd ed., rev.* (Harvard, 1980): 20–46.

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. Phil. Qtly.* 19: 289–298.

Rapaport, W.J. (1986a), "Searle's Experiments with Thought", *Phil. Sci.* 53: 271–279.

Rapaport, W.J. (1986b), "Logical Foundations for Belief Representation," *Cog. Sci.* 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. Experimental and Theoretical AI* 10: 403–419.

Rapaport, W.J. (2000), "How to Pass a Turing Test", *J. Logic, Lang., and Info.*, 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 Int'l. Conf. Cog. Sci./7th Australasian Soc. Cog. Sci. Conf.* (Sydney: U. New South Wales) 2: 547–552.

Rapaport, W.J. (2005), "In Defense of Contextual Vocabulary Acquisition", in A. Dey, B. Kokinov, D. Leake, & R. Turner (eds.), *Modeling and Using Context* (Berlin: Springer-Verlag Lecture Notes in AI 3554): 396–409.

Rapaport, W.J. (2006a), "The Turing Test", in K. Brown (ed.), *Encyc. Lang. and Ling., 2nd Ed.* 13: 51–159 (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. (2007), "Essential SNePS Readings" [http://www.cse.buffalo.edu/∼rapaport/snepsrdgs.html].

Rapaport, W.J. (compiler) (2008), "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 Ł. Iwańska & S.C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation* (Menlo Park, CA/Cambridge, MA: AAAI/MIT Press): 347–375; errata at [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", *J. Experimental and Theoretical AI*: 19: 1-17.

Rapaport, WJ.; & Shapiro, S.C. (1984), "Quasi-Indexical Reference in Propositional Semantic Networks," *Proc. 10th Int'l Conf. Comp. Ling.* (Morristown, NJ: Assoc. Comp. Ling.): 65–70.

Rapaport, W.J., & Shapiro, S.C. (1995), "Cognition and Fiction", in J.F. Duchan, G.A. Bruder, & L.E. Hewitt (eds.), *Deixis in Narrative* (Hillsdale, NJ: Erlbaum): 107–128.

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* (Cambridge, MA: MIT): 11–25.

Rayner, K., & Well, A.D. (1996), "Effects of Contextual Constraint in Eye Movements in Reading", *Psychonomic Bull. & Rev.* 3: 504–509.

Reber, A.S. (1989), "Implicit Learning and Tacit Knowledge", *J. Experimental Psych.: 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* (Hillsdale, NJ: Erlbaum): 27–48.

Rumelhart, D.E. (1980), "Schemata: The Building Blocks of Cognition", in R.J. Spiro et al. (eds.), *Theoretical Issues in Reading Comprehension* (Hillsdale, NJ: 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: Int'l. Reading Assn.).

Russell, B. (1905), "On Denoting", reprinted in B. Russell, *Logic and Knowledge: Essays 1901–1950*, R.C. Marsh (ed.), (New York: Capricorn, 1956): 39–56.

Russell, Bertrand (1918), "The Philosophy of Logical Atomism", reprinted in *Logic and Knowledge: Essays 1901–1950*, R.C. Marsh (ed.) (New York: Capricorn, 1956): 177–281.

Saussure, F. de (1959), *Course in General Linguistics*, C. Bally, A. Sechehaye, & A. Reidlinger (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 Understanding: Teaching from the Perspective of Artificial Intelligence* (Hillsdale, NJ: Erlbaum).

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

Schatz, E.K., & Baldwin, R.S. (1986), "Context Clues Are Unreliable Predictors of Word Meanings", *Reading Res. Qtly.* 21: 439–453.

Schwartz, R.M. (1988), "Learning to Learn Vocabulary in Content Area Textbooks", *J. 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", *Psych. Bull.* 115: 163–196.

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

Shapiro, S.C. (1982), "Generalized Augmented Transition Network Grammars for Generation from Semantic Networks", *Amer. J. Comp. Ling.* 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.*, *Lecture Notes in AI 390* (Berlin: Springer-Verlag): 362–380.

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

Shapiro, S.C. (1992), "Artificial Intelligence", in S.C. Shapiro (ed.), *Encyc. AI, 2nd ed.* (New York: Wiley): 54–57.

Shapiro, S.C. (2001), "Computer Science: The Study of Procedures", [http://www.cse.buffalo.edu/~shapiro/Papers/whatiscs.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* (New York: Springer-Verlag): 262–315.

Shapiro, S.C., & Rapaport, W.J. (1992), "The SNePS Family," *Computers and 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, G.A. Bruder, & L.E. Hewitt (eds.), *Deixis in Narrative* (Hillsdale, NJ: Erlbaum): 79–105.

Shapiro, S.C., & the SNePS Implementation Group (2007), "SNePS 2.7 User's Manual" [http://www.cse.buffalo.edu/sneps/Manuals/manual27.pdf].

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

Share, D.L. (1999), "Phonological Recoding and Orthographic Learning", *J. Experimental Child Psych.* 72: 95–129.

Shieber, Stuart M. (ed.) (2004). *The Turing Test: Verbal Behavior as the Hallmark of Intelligence* (Cambridge, MA: MIT Press).

Simon, H.S. (1996), "Computational Theories of Cognition", in W. O'Donohue & R.F. Kitchener (eds.), *Philosophy of Psychology* (Sage): 160–172.

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

Singer, M.; Revlin, R.; & Halldorson, M. (1990), "Bridging-Inferences and Enthymemes", in Graesser & Bower 1990: 35–51.

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", *Artif. Intell.* 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 U. Press).

Stanovich, K.E. (1986), "Matthew Effects in Reading", *Reading Res. Qtly.* 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* (Hillsdale, NJ: Erlbaum): 89–105.

Sternberg, R.J., & Powell, J.S. (1983), "Comprehending Verbal Comprehension", *Amer. 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* (New York: Academic): 121–143.

Suh, S., & Trabasso, T. (1993), "Inferences during Reading", *J. Memory and Lang.* 32: 279–300.

Swanborn, M.S.L., & De Glopper, K. (1999), "Incidental Word Learning while Reading", *Rev. Ed. Res.* 69: 261–285.

Szabó, Z.G. (2007), "Compositionality", in E.N. Zalta (ed.), *Stanford Encyc. Phil. (Spring)* [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 Qtly.* 30: 415–433.

Turing, A.M. (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", *J. Verbal Learning and Verbal Behav.* 20: 386–399.

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

Webster, M.A. (2004), "Adaptation and the Phenomenology of Perception", unpublished talk, SUNY Buffalo Center for Cog. Sci. [http://wings.buffalo.edu/cogsci/Activities/Colloquium/CLLQf04/2004fall.htm#webster04]

Werner, H., & Kaplan, E. (1952), "The Acquisition of Word Meanings", *Monographs, Soc. Res. Child Development* 15 (1950) 51(1).

Wesche, M., & Paribakht, T.S. (eds.) (1999), "Incidental L2 Vocabulary Acquisition", *Studies in 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", Ph.D. diss. (Buffalo: SUNY Buffalo Dept. 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), "Invented Names, Hidden Distortions in Social Science", *Chronicle Higher Ed.* (May 30): B13–B14.

Woods, W.A. (1970), "Transition Network Grammars for Natural Language Analysis", *Comm. ACM* 13: 591–606.

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[a]
x is a P i.e., NP_{indiv} is an NP_{common}[b] 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[c]
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[d]
x is a synonym of y	x and y are synonyms

Table 1:

---

[a]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.

[b]I.e., a sentence consisting of a noun phrase representing an individual, followed by 'is a', followed by a common-noun phrase.

[c]For more information on the possessive "x is y's R", see Rapaport 2006b.

[d]Shapiro & Rapaport 1987.