

# In Defense of Contextual Vocabulary Acquisition

## How to Do Things with Words in Context

William J. Rapaport

Departments of Computer Science and Engineering and Philosophy;  
Center for Cognitive Science,  
State University of New York at Buffalo, Buffalo, NY 14260-2000, USA  
+1-716-645-3180x112; fax +1-716-645-3464  
rapaport@cse.buffalo.edu  
[http://www.cse.buffalo.edu/~sim\\$rapaport/CVA/](http://www.cse.buffalo.edu/~sim$rapaport/CVA/)

**Abstract.** Researchers in “contextual vocabulary acquisition” differ over the kinds of context involved in vocabulary learning, and the methods and benefits thereof. This paper presents a computational theory of contextual vocabulary acquisition, identifies the relevant notion of context, exhibits the assumptions behind some classic objections, and defends our theory against these objections.

## 1 A Computational Theory of Contextual Vocabulary Acquisition

*Contextual vocabulary acquisition* (CVA) is the deliberate acquisition of a meaning for a word in a text by reasoning from context, where “context” *includes*: (1) the reader’s “internalization” of the surrounding text, i.e., the reader’s “mental model” of the word’s “*textual* context” (hereafter, “co-text” [3]) integrated with (2) the reader’s prior knowledge (PK), but it *excludes* (3) external sources such as dictionaries or people. CVA is what you do when you come across an unfamiliar word in your reading, realize that you don’t know what it means, decide that you need to know what it means in order to understand the passage, but there is no one around to ask, and it is not in the dictionary (or you are too lazy to look it up). In such a case, you can try to figure out its meaning “from context”, i.e., from clues in the co-text together with your prior knowledge.

Our computational theory of CVA—implemented in a the SNePS knowledge representation and reasoning system [28]—begins with a stored knowledge base containing SNePS representations of relevant PK, inputs SNePS representations of a passage containing an unfamiliar word, and draws inferences from these two (integrated) information sources. When asked to define the word, definition algorithms deductively search the resulting network for information of the sort that might be found in a dictionary definition, outputting a definition frame 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 [6–8,20,23,24].

We are investigating ways to make our system more robust, to embed it in a natural-language-processing system, and to incorporate morphological information. Our research group, including reading educators, is also applying our methods to the develop-

ment of what we hope will be a better pedagogical curriculum than the current state of the art for teaching CVA.

To this end, we have been studying the CVA literature from a variety of disciplines that, generally speaking, seem to ignore each other's literature (including computational linguistics, reading education, second-language education, and psychology [22]). Two often-cited papers by reading scientists [2,26] have claimed that not only are certain contexts less than useful for doing CVA, but that most "natural" (as opposed to artificial) contexts are not helpful at all. Their arguments make several assumptions inconsistent with our computational theory. Thus, their objections do not apply to us.

## 2 Are All Contexts Created Equal?

### 2.1 The Role of Prior Knowledge

Beck et al.'s [2], subtitled "All Contexts Are Not Created Equal", claims that "it is not true that every context is an appropriate or effective instructional means for vocabulary development". They begin by pointing out that "the context that surrounds a word in text can give *clues to the word's meaning*" (my italics). But a passage is not a clue without some other information to interpret it as a clue. Therefore **(A1) Textual clues must be supplemented with other information in order for a meaning to be computed.** *This supplemental information must come from the reader's PK.* Such PK (which need not be true) might include general "world" or cultural knowledge, "commonsense" knowledge, specialized "domain" knowledge, and perhaps the "background" knowledge the author assumes the reader will have. However, not all of the reader's PK may be consciously available at the time of reading, and each reader will bring to bear upon his or her interpretation of the text *idiosyncratic* PK [10,12].

I will use 'co-text' to refer to the text surrounding an unfamiliar word, reserving 'context' or 'wide context' to refer to the reader's available PK "integrated" with the reader's "internalization" (or "mental model") of the co-text. Its integration with the reader's PK involves belief revision: New beliefs would be inferred as conclusions from arguments in which at least one premise comes from the internalized co-text and at least one premise comes from PK. Typically, withdrawn beliefs are PK beliefs inconsistent with co-text propositions [20]. The "context" that the reader uses to compute a word's meaning is not just the *co-text* but this *wider* context.

The reader's internalization of the text involves some interpretation (e.g., resolving pronoun anaphora) or the immediate and unconscious drawing of an inference (e.g., that 'he' refers back to a male or that 'John' is a proper name typically referring to a male human) [10]. Consider the following natural passage (my italics): "The archives of the medical department of Lourdes are filled with *dossiers* that detail well-authenticated cases of what are termed miraculous healings" [17]. 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 might not be sensitive to this distinction. 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 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. (We have anecdotal evidence that at least some readers of this passage interpret it in the latter way.)

Even a common word can mean different things to different people: Something that looks like a sofa but seats only one is a ‘sofa’ in Indian English but a ‘chair’ in American English. Thus, two fluent English speakers might interpret a passage containing the word ‘sofa’ differently: The text is the same in both cases, but the readers’ *internalized* texts will differ. It can also involve simple misreading: I 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 I never heard of this local bakery? Why might they be delivering a cake? So I re-read the truck’s sign more carefully. It actually said, “Mills Welding and Specialty Gases”! A related modifying influence stems from reading difficulties that might circumscribe the amount of co-text that the reader can understand and therefore integrate into his or her mental model [30].

## 2.2 Do Words Have Unique Meanings?

The assumption—inconsistent with our theory—underlying [2]’s use of the phrase ‘the word’s meaning’ is **(A2) A word has a unique meaning**. The definite description ‘*the* word’s meaning’ or ‘*the* meaning of a word’ is ubiquitous but worth avoiding, for it incorrectly suggests that a word has a unique meaning. Perhaps what’s normally intended by this phrase is “the meaning of a word *in the present context*”: “[C]ontext always *determines* the meaning of a word, it does not always *reveal* it” (Deighton, cited in [26]). But it follows from our observations about **(A1)** that the reader will supplement the co-text with idiosyncratic PK; consequently, each reader will interpret the word slightly differently. Deighton is still essentially correct: Wide context determines *a* meaning for the word, though it requires further processing to reveal that meaning.

The need for further processing underlies [14]’s observation that we don’t store definitions, even of words we understand. It also undercuts pedagogical strategies for CVA that instruct the reader merely to “guess” the meaning [4]. Nation [19] even boasts that his guessing strategy “does not draw on background content knowledge” since “linguistic clues will be present in every context, background clues will not”. But background knowledge (PK) is essential and always used; it is unavoidable, even in Nation’s own strategy: Where he says “Guess” (the entire step 4 in his 5-part strategy!), 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 PK [20,29].

### 2.3 Do Words Have Correct Meanings?

A closely related assumption that many authors make is (A3) **There is such a thing as “the correct meaning of a word”**. “[E]ven 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” [2; my italics].

Perhaps what is meant by the “correct” meaning is that there is a certain meaning that *the author* intended. But if we are concerned with a word’s meaning *as determined by* the reader’s internalized co-text integrated with the reader’s PK, then 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, rather than *the* meaning *of* the word (i.e., the reader *gives* or *assigns* a meaning *to* the word).

If the meaning that the reader computes *is* the intended one, so much the better. If not, has the reader then *misunderstood* the text? Misunderstanding is not necessarily a bad thing: If no one ever understood texts differently from other readers or from the author’s intended meaning, 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 [21]. This is the mechanism that makes conversation and the exchange of information possible [25]. The important question is not whether a reader can compute *the correct* meaning of a word, but whether the reader can compute *a* meaning for the word *that is sufficient to enable him or her to understand the text*. The reader need not understand the text “perfectly”, but merely *well enough to continue reading*. We don’t normally have, *nor do we need*, full, correct definitions of the words we understand [14].

Consider the following passage:<sup>1</sup> “All chances for agreement were now gone, and compromise would now be impossible; in short, an *impasse* had been reached” [5]. A reader might compute a meaning for ‘impasse’ from this text thusly: A compromise is an agreement. If all chances for agreement are gone, then agreement is impossible. So both conjuncts of the first clause say almost the same thing. ‘In short’ is a clue that what follows means almost the same as what precedes it. So, to say that an impasse has been reached is to say that agreement is impossible. And (perhaps with a bit more plausible PK) that means that an impasse is a *disagreement*. *Is it?* At least one dictionary defines it simply as a “deadlock”. Suppose that “deadlock” is “the correct meaning” of ‘impasse’. 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 not “correctly” understood the word (though, in the case of this particular bit of CVA, she has surely computed *a* very plausible meaning). (2) The reader sees the word again in a context in which “disagreement” is a plausible meaning. Since her PK 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 “deadlock”, not

<sup>1</sup> From an article detailing teachable contextual clues for CVA; so this might be a “pedagogical”, not a “natural”, passage. Our project reports are at [<http://www.cse.buffalo.edu/~rapaport/CVA/>].

“disagreement”, is the “best” meaning. E.g., she might read a 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: (3a) The reader might decide that this occurrence of ‘impasse’ could not possibly mean “disagreement”. Again, there are two possibilities: (3a-i) She decides that she must have been wrong about ‘impasse’ meaning “disagreement”, and she now comes to believe (say) that it means “deadlock”. (3a-ii) She decides that ‘impasse’ is polysemous, and that “deadlock” is a second meaning. (Cf. the polysemous verb ‘to dress’ [23]; a reader might firmly believe that to dress is to put clothes on but, from co-texts such as “King Claudas dressed his spear before battle”, infers that to dress is *also* to prepare for battle.) (3b) Or the reader might try to reconcile the two possible meanings, perhaps by viewing deadlocks as disagreements, if only metaphorically [1].

#### 2.4 Two Kinds of Textual Context

Beck et al. are interested in using *co-text* to help *teach* “the meaning” of the word. We, however, are interested in using *wide* context to help compute a meaning for an unfamiliar word, for the purpose of *understanding the passage* containing it. These two interests don’t always coincide, especially if the former includes as one of its goals the reader’s ability to *use* the word. From the fact that a given co-text might not clearly convey a word’s “correct” meaning, it does not follow that a useful meaning cannot be computed from it (especially since the wider context from which a meaning is computed includes the reader’s PK and is not therefore restricted to the co-text). Some co-texts certainly provide more clues than others. But should all CVA be spurned because of less-helpful co-texts?

Their classification divides all co-texts into *pedagogical* and *natural*. The former are “specifically designed for teaching designated unknown words”. It will be of interest later that the only example they give of a pedagogical co-text is for a *verb*: “All the students made very good grades on the tests, so their teacher *commended* them for doing so well.”

By contrast, “the author of a natural context does not intend to convey *the meaning of a word*” (my italics). Note the assumptions about unique, correct meanings at work. In contrast, and following Deighton (§2.2, above), the author of a natural co-text *does*—no doubt, unintentionally—convey *a* meaning for the word in question. And that meaning is the only one that a reader might be expected to compute. [2] goes on to observe that natural “contexts will not necessarily provide *appropriate* cues to the meaning of a particular word” (my italics). This does not mean that no cues (or clues) are provided. It may well be that clues *are* provided for *a* meaning that helps the reader understand the passage. Note that the pedagogical-natural distinction may ultimately be hard to maintain: A passage produced for pedagogical purposes by one researcher might be taken as “natural” by another (see §2.6, below).

#### 2.5 Four Kinds of (Natural) Co-texts

**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” (my italics). Some co-texts may indeed be misdirective. But [2]’s sole example does not inspire confidence: “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.” Granted, a reader might incorrectly decide from this that ‘grudgingly’ meant something like “admiringly”. But there are three problems with this example: (1) There is no evidence that this co-text is natural. This is minor; many such allegedly misdirective co-texts could be found “in nature”. (2) If it is natural, it would be nice to see more of it. Many CVA researchers assume that **(A4) co-texts have a fixed, usually small size**. But there might be other clues, preceding or following this short co-text, that would rule out “admiringly”. Perhaps we know or could infer from other 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. (3) Most significantly, ‘grudgingly’ is an adverb. Another assumption is **(A5) All words are equally easy (or difficult) to learn**. But adverbs, adjectives, and other modifiers are notoriously hard cases for CVA and for first-language learning [11].

Thus, the evidence provided for the existence of misdirective co-texts is weak, primarily since there should be *no* limit on the size of a co-text (see §3.2, below) and since 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. Certainly a reader’s PK (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.

Beck et al. conclude, “[I]ncorrect conclusions about word meaning are likely to be drawn” from misdirective co-texts. This assumes **(A6) Only one co-text can be used to compute a meaning for a word**. Granted, if a word only occurs once, in the most grievous of misdirective co-texts, then it is quite likely that a reader would “draw an incorrect conclusion”. But, in such a case, it does not matter if the reader even 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 any initial hypothesis about what it might mean.

The task of CVA is hypothesis generation and testing, a fundamentally scientific task of developing a theory about a word’s (possible) meaning. It is not mere guessing. It is like detective work: finding clues to determine, not “who done it”, but “what it means”. And, like all hypotheses, theories, and conclusions drawn from circumstantial evidence, it is susceptible to revision when more evidence is found.

Admittedly, all of this assumes that the reader is consciously aware of the unfamiliar word and notes its unfamiliarity. It also assumes that the reader remembers the word and its hypothesized meaning (if any) between encounters. None of these further assumptions are, unfortunately, necessarily the case.

**Nondirective Co-texts.** “[N]ondirective contexts, . . . seem to be of *no* assistance in directing the reader toward any particular meaning for a word” (my italics). Here is [2]’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.” 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). I suggested that the reader could ignore a single unfamiliar word in a misdirective text. The same is true of a non-directive text. 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 PK includes the author’s intended meaning for that word. As a literary conceit, it might be excusable; in expository writing, it would not be.

**Syntactic Manipulation.** All co-texts (even misdirective and non-directive) are capable of yielding a clue. The technique for squeezing a clue out of any co-text is to syntactically manipulate it to make the unfamiliar word its focus, 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. For example, from the above “misdirective” text, we could infer that, whatever else ‘grudgingly’ might mean, it could be defined (if 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). I put ‘apparently’ in parentheses, because some readers, depending on their PK, might realize that sometimes praise can be given reluctantly or ironically, and such readers might hypothesize that ‘grudgingly’ is that kind of way of praising. 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 ([19] makes similar remarks).

**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”. E.g., “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.” Note that this adjective is contrasted with Stan’s attitude. From a contrast, much can be inferred. In our research, several adjectives that we have computed meanings for occur in such contrastive co-texts: “Unlike his brothers, who were noisy, outgoing, and very talkative, Fred was quite *taciturn*” [5] (though this is probably not a natural co-text, or else it is a “directive” co-text.)

**Directive Co-texts.** Their fourth category is “directive contexts, which seem likely to lead the student to a specific, correct meaning for a word”. 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*?’ ” Again, it’s not clear whether this is a natural co-text. 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!)

## 2.6 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”. 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.) (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 assumption: **(A7) CVA “comes naturally”, hence needs no training.** Our project, by contrast, is not focused on incidental CVA, but on deliberate CVA, carefully taught and practiced.

(4) Another problem arises from the next assumption: **(A8) Cloze-like tasks are a form of CVA.** A “cloze-like” task involves replacing certain words in a passage with blanks to be filled in. This is not CVA. A serious methodological difficulty faces all CVA researchers: If you want to find out if a subject can compute 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 who knows which words. But then those who don’t know the test words will have seen them at least once before (during the pretest), 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: 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 related, yet distinctly different.<sup>2</sup> Schatz & Baldwin [26] 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, the reader is invited to guess (rather than compute), and there is a unique, correct answer, whereas, in CVA as we see it, the goal is to compute a meaning sufficient for understanding the passage.

I have no clever solution to this methodological problem. My 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 not have a single-word counterpart in English, but that in any case the subject’s job is to compute what it might mean, not necessarily find an English synonym, exact or inexact.

---

<sup>2</sup> I am grateful to my co-researcher, Michael Kibby, for this insight.

## 2.7 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”. However, “Only one subject could identify any word in the misdirective category”. This is significant because it suggests that CVA *can* be done even with misdirective co-texts, supporting *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”. An assumption underlying this is: **(A9) CVA can be of help only in vocabulary acquisition.** But another potential benefit far outweighs this: 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 to computing a meaning; application of relevant PK; application of reasoning for the purpose of extracting information from the text.

## 3 Are Context Clues Unreliable Predictors of Word Meanings?

### 3.1 Schatz & Baldwin’s Argument

Schatz & Baldwin [26] 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”.

In summarizing the then-current state of the art, they 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” (my italics). Given their rhetoric, the underlying assumption here appears to be: **(A10) CVA is not an efficient mechanism for inferring word meanings.** Their argument seems to be, roughly, that *co-text* can’t help you figure out “the” correct meaning of an unfamiliar word; therefore, CVA is not “an effective strategy for inferring word meanings”. In contrast, I am arguing that *wide* context *can* help you figure out *a* meaning for an unfamiliar word; therefore, CVA *is* an effective strategy for inferring (better: computing) word meanings. Insofar as the *purpose* of CVA is thought of as getting “the correct meaning”, it is ineffective. But insofar as 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 “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*” (my italics). This

could only be the case if CVA were completely unconscious and immediate, as if one could read a passage with an unfamiliar word and instantaneously come to know what it means. (This *may* hold for “incidental” CVA [18], but not for “deliberate” CVA.) In contrast, our theory requires interruption—not to access external sources—but for conscious, deliberate analysis of the passage. 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. With the exception of learner’s dictionaries designed primarily for ESL audiences, most dictionaries are notoriously difficult to use and their definitions notoriously difficult to interpret [16]. More importantly, *CVA needs to be applied to the task of understanding a dictionary definition itself*, which is, after all, merely one more co-text containing the unfamiliar word [27]. Indeed, CVA is the base case of a recursion one of whose recursive clauses is “look it up in a dictionary”.

### 3.2 Schatz & Baldwin’s Methodology

**Nouns and Verbs vs. Modifiers.** Schatz & Baldwin offer several experiments to support their claims. As with [2]’s experiments, there are a number of apparent problems with their methodology. Their first experiment took 25 “natural” passages from novels, selected according to an algorithm that randomly produced passages containing low-frequency words. But consider some of their words: ‘cogently’, ‘cozened’, ‘ignominiously’, ‘imperious’, ‘inexorable’, ‘perambulating’, ‘recondite’, ‘salient’. Note that 4 are adjectives, 2 are adverbs, 1 is a verb (‘cozened’), and 1 (‘perambulating’) might be either a verb or an 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. If they are, then fully 75% of the unfamiliar words are modifiers, known to be among the most difficult of words to learn. Their 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% modifiers, 27% nouns, and 6% verbs (not counting ‘perambulating’). Of these, two of the nouns (‘dearth’, ‘ameliorating’) are examples of words occurring in “facilitative” co-texts. 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? They admit that “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”. A representative sample of co-texts? 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 [2]’s categories with nouns, verbs, adjectives, and adverbs in each such co-text (i.e., 16 possible types of co-text)? Schatz & Baldwin’s (and [2]’s) results may say more about the difficulty of learning meanings for modifiers than they do about weaknesses of contexts.

**CVA vs. WSD.** Moreover, in two of their experiments, subjects were *not* involved in the task of CVA. Rather, they were doing a related—but distinct—task known as “word-

sense disambiguation” (WSD [13]). The CVA task is to figure out a word’s meaning “from scratch”. The WSD task is to choose a meaning for a (typically polysemous) word from a list of possible meanings for the word in different contexts. In [26]’s experiment, the subjects only 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 possible meanings fit better; no real CVA was needed.

In the third experiment, real CVA was being tested. However, (A3) raises its head: “we were interested only in full denotative meanings or accurate synonyms”. There is no reason to believe or 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 is a useful tool for vocabulary acquisition and general reading comprehension.

**Space and Time Limits.** The smaller the co-text, the less chance there is of figuring out a meaning, because there will be a minimum of textual clues. The larger the co-text, the greater the chance, because a large enough co-text might actually include a definition of the word! (Recall 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 and work “outwards” to preceding and succeeding passages, until enough co-text is provided to enable successful CVA. ‘Successful’ only means being able to compute *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. In contrast, [26] arbitrarily limited their co-text size to only 3 surrounding sentences. 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.

Also, [26] observes that “All students finished in the allotted time”. But real-life CVA has no time limits (other than self-imposed ones). CVA might extend over a long period of time, as different texts are read.

**Teaching CVA.** 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”. Their finding “that students either could not or chose not to use context to infer the meanings of unknown words” ignores the possibilities that the students did not know that they *could* use context or that they did not know *how* to. Granted, “incidental” (unconscious) CVA is something that we all do; there appears to be no other explanation for how we learn most of our vocabulary [18]. 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, their conclusion that “context is an ineffective or little-used strategy for helping students infer the meanings of low-frequency words” might only be true for untrained readers. It remains an open question whether proper training in CVA can make it effective and can add it to the reader’s arsenal of techniques for improving reading comprehension (though there is some positive evidence [9,15]). [26] disagrees: “[I]f 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.” (A7) is at work again. But 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 strongly suggests that students need to, and can, be educated on these matters.

### 3.3 Three Questions About CVA

In their general-discussion section, [26] raises three questions: (1) “Do traditional context clues occur with sufficient frequency to justify them as a major element of reading instruction?” This is irrelevant *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. CVA can foster improved reading comprehension, but more research is needed. Traditional context clues do occur and—augmented by the reader’s PK and training in CVA techniques for developing revisable hypotheses about an unfamiliar word’s meaning—are justified as a major element of reading instruction.

(2) “Does context *usually* provide accurate clues to the denotations and connotations of low-frequency words?” This is also irrelevant under our conception of CVA: We are not interested in “accuracy”. 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 [26] (or [2]) seem to have in mind. 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”? Such words are always amenable to yielding at least some information about their meaning, as discussed in §2.5, above.

## 4 Conclusions: A Positive Theory of Computational CVA

Progress is often made by questioning assumptions. This essay has questioned the assumptions underlying [2]’s and [26]’s arguments and experiments challenging CVA. Their papers are best read as asserting that, *given those assumptions*, CVA is not as beneficial as some researchers claim it is. We conclude by presenting our theory’s contrasting beliefs. (Details are in [23,24].)

**(B1)** Every co-text  $C$  can give some clue (even minimally) to a word  $w$ ’s meaning (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 reader  $R$ ’s accessible PK, which may be time-dependent. None of the meanings  $R$  computes for  $w$  is necessarily “the” meaning (in either a dictionary sense or that of a reading teacher).

**(B2)**  $w$ ’s co-text gives clues to  $w$ ’s meaning that must be supplemented by the reader’s PK in order for a meaning to be computed. There is no such things as “misdirective”, “non-directive”, “general”, or “directive” co-texts. A co-text’s value depends on the reader’s PK and ability to use clues and PK together.

**(B3)** CVA is distinct from cloze-like tasks.

**(B4)** Co-texts can be as small as a phrase or as large as an entire book, with no arbitrary space or time limits.

**(B5)** Many co-texts may be required before CVA can “asymptotically” approach a “stable” meaning for a word.

**(B6)** A word does not have a unique meaning, even in directive and pedagogical co-texts.

**(B7)** A word does not have a (single) correct meaning, not even in directive and pedagogical co-texts. Nor does it *need* a correct meaning in order for a reader to be able to understand it (in context). Even a familiar and well-known word can acquire a new meaning in a new co-text. In fact, each new *C* and each new *R* can yield a new meaning, so meanings are continually being extended (as when words are used metaphorically [4]).

**(B8)** Some words are harder to compute meanings for than others (e.g., nouns are easiest).

**(B9)** CVA is an efficient method for inferring word meanings.

**(B10)** CVA can improve general reading comprehension.

**(B11)** CVA can (and should) be taught.<sup>3</sup>

## References

1. Budiu, R. & Anderson, J.R. (2001), “Word Learning in Context: Metaphors and Neologisms”, *Tech. Rep. CMU-CS-01-147* (Carnegie Mellon Univ., School of Comp. Sci.).
2. Beck, I.L.; McKeown, M.G.; & McCaslin, E.S. (1983), “Vocabulary Development: All Contexts Are Not Created Equal”, *Elementary School Journal* 83: 177–181.
3. Brown, G., & Yule, G. (1983), *Discourse Analysis* (Cambridge Univ. Press).
4. Clarke, D.F., & Nation, I.S.P. (1980), “Guessing the Meanings of Words from Context”, *System* 8: 211–220.
5. Dulin, K.L. (1970), “Using Context Clues in Word Recognition and Comprehension”, *Reading Teacher* 23: 440–445, 469.
6. Ehrlich, K. (1995), “Automatic Vocabulary Expansion through Narrative Context”, *Tech. Rep. 95-09* (SUNY Buffalo Comp. Sci.).
7. Ehrlich, K., & Rapaport, W.J. (1997), “A Computational Theory of Vocabulary Expansion”, *Proc. 19th Annual Conf. Cog. Sci. Soc.* (Erlbaum): 205–210.
8. Ehrlich, K., & Rapaport, W.J. (2004), “A Cycle of Learning: Human and Artificial Contextual Vocabulary Acquisition”, *Proc. 26th Annual Conf. Cog. Sci. Soc.* (Erlbaum, 2005): 1555.
9. Fukkink, R.G., & De Glopper, K. (1998), “Effects of Instruction in Deriving Word Meaning from Context”, *Rev. Ed. Res.* 68: 450–458.
10. Garnham, A., & Oakhill, J. (1990), “Mental Models as Contexts for Interpreting Texts”, *J. Semantics* 7: 379–393.
11. Gentner, D. (1982), “Why Nouns Are Learned before Verbs”, in S.A. Kuczaj (ed.), *Language Development: Vol. 2* (Erlbaum): 301–334.
12. Hobbs, J.R. (1990), *Literature and Cognition* (Stanford, CA: CSLI).
13. Ide, N.M., & Veronis, J. (eds.) (1998), Word Sense Disambiguation, *Comp. Ling.* 24(1).
14. Johnson-Laird, P.N. (1987), “The Mental Representation of the Meanings of Words”, *Cogn.* 25: 189–211.

<sup>3</sup> I am grateful to Albert Goldfain, Michael W. Kibby, Jean-Pierre Koenig, Shakthi Poornima, Stuart C. Shapiro, Karen M. Wieland, and the SNePS Research Group. A longer version of this paper is at [<http://www.cse.buffalo.edu/~rapaport/Papers/paris.pdf>].

15. Kuhn, M.R., & Stahl, S.A. (1998), "Teaching Children to Learn Word Meanings from Context", *J. Lit. Res.* 30: 119–138.
16. Miller, G.A. (1986), "Dictionaries in the Mind", *Lang. & Cog. Procs.* 1: 171–185.
17. Murphy, J. (2000), *The Power of Your Subconscious Mind, Rev. Ed.* (Bantam).
18. Nagy, W.E.; Herman, P.A.; & Anderson, R.C. (1985), "Learning Words from Context", *Reading Res. Qlty.* 20: 233–253.
19. Nation, I.S.P. (2001), *Learning Vocabulary in Another Language* (Cambridge Univ. Press).
20. Rapaport, W.J. (2003a), "What Is the 'Context' for Contextual Vocabulary Acquisition?", in P.P. Slezak (ed.), *Proc. ICCS/ASCS-2003*, Vol. 2: 547–552.
21. Rapaport, W.J. (2003b), "What Did You Mean by That? Misunderstanding, Negotiation, and Syntactic Semantics", *Minds & Machines* 13: 397–427.
22. Rapaport, W.J. (2004), "Bibliography of Theories of Contextual Vocabulary Acquisition", [<http://www.cse.buffalo.edu/~rapaport/refs-vocab.html>].
23. Rapaport, W.J., & Ehrlich, K. (2000), "A Computational Theory of Vocabulary Acquisition", in L.M. Iwańska & S.C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation* (AAAI/MIT Press): 347–375.
24. Rapaport, W.J., & Kibby, M.W. (2002), "Contextual Vocabulary Acquisition: A Computational Theory and Educational Curriculum", in N. Callaos et al. (eds.), *Proc. SCI-2002*, Vol. II: 261–266.
25. Russell, B. (1918), "The Philosophy of Logical Atomism", in *Logic and Knowledge* (Capricorn, 1956): 177–281.
26. Schatz, E.K., & Baldwin, R.S. (1986), "Context Clues Are Unreliable Predictors of Word Meanings", *Reading Res. Qlty.* 21: 439–453.
27. Schwartz, R.M. (1988), "Learning to Learn Vocabulary in Content Area Textbooks", *J. Reading*: 108–118.
28. Shapiro, S.C., & Rapaport, W.J. (1995), "An Introduction to a Computational Reader of Narrative," in J.F. Duchan et al. (eds.), *Deixis in Narrative* (Erlbaum): 79–105.
29. Singer, M., et al. (1990), "Bridging-Inferences and Enthymemes", in A.C. Graesser & G.H. Bower (eds.), *Inferences and Text Comprehension* (Academic).
30. Stanovich, K.E. (1986), "Matthew Effects in Reading", *Reading Res. Qlty.* 21: 360–407.