Abstract

Elman (2009) proposed that the traditional role of the mental lexicon in language processing can largely be replaced by a theoretical model of schematic event knowledge founded on dynamic context-dependent variables. We evaluate Elman’s approach and propose an alternative view, based on dual coding theory and evidence that modality-specific cognitive representations contribute strongly to word meaning and language performance across diverse contexts which also have effects predictable from dual coding theory.

Keywords: Lexicon; Context; Imagery; Event representation; World knowledge

1. Introduction

Elman’s (2009) theory of lexical knowledge without a mental lexicon is based on analyses of (a) problems with the standard view of the mental lexicon; (b) the roles of world knowledge and linguistic knowledge in comprehension of event descriptions; (c) experimental evidence that word meanings are heavily dependent on their sentence contexts; and (d) the possibility of computational modeling of his approach. We re-analyze these issues from the perspective of dual coding theory (DCT) and find agreements and disagreements with different aspects of Elman’s position. We especially find common ground in his emphasis on the importance of world knowledge in comprehension of event descriptions, but not his ultimate schema interpretation of how that knowledge is represented. We agree too on the importance of context in determining word meaning, but we attach much more weight than he does to stable semantic properties of language units. Finally, while respecting his goal of
providing a formal computational theory to explain event knowledge, we find that this goal remains unsure and elusive. In contrast, noncomputational DCT has successfully explained and predicted relevant phenomena and can be easily extended to the specific language domains addressed by Elman. We elaborate on these comparisons after summarizing the DCT approach to the focal issues.

2. Mental “lexicon,” context, and DCT

Elman (2009) emphasized knowledge complexity and context dependence as major problems associated with current theories of the mental lexicon. This applies particularly to the standard view that word information is stored mentally in a single, abstract (lemma) form applicable to different modalities of language (e.g., auditory, visual, motor) along with their semantic, syntactic, and pragmatic properties. The inclusion of pragmatic properties particularly implies that contextual influences on language performance arise from the language user’s world knowledge as well as from the language context.

The mental “lexicon” coupled with corresponding nonlinguistic representations are the core of DCT structures and processes (e.g., Paivio, 1971, 1986, 2007; Sadoski & Paivio, 2001, 2004). However, these representations are modality-specific and embodied rather than abstract as they are in the standard theories, although abstract phenomena (e.g., abstract language) can be handled by the DCT assumptions. Initially, the units were simply called verbal and nonverbal (imaginal) internal representations that vary in sensorimotor modality (Paivio, 1971; pp. 54–56). Subsequently, the verbal and nonverbal units were named logogens and imagens, respectively (Paivio, 1978).

Logogen was adapted from Morton’s (1979) word recognition model in which the logogen was interpreted as a multimodal concept that includes auditory and visual logogens as well as input and output logogens. In DCT (e.g., Paivio, 1986; Sadoski & Paivio, 2001) the concept expanded to include auditory, visual, haptic, and motor logogens, as well as separate logogen systems for the different languages of bilinguals.

Logogens of all modalities are hierarchical sequential structures of increasing length, from phonemes (or letters) to syllables, conventional words, fixed phrases, idioms, sentences, and longer discourse units—anything learned and remembered as an integrated sequential language unit.

The DCT imagens are mental representations that give rise to conscious imagery and are involved as well in recognition, memory, language, and other functional domains. Like logogens, imagens come in different modalities—visual, auditory, haptic, and motor. They also are hierarchically organized but, in the case of visual imagens in particular, the hierarchy consists of spatially nested sets—pupils within eyes within faces, rooms within houses within larger scenes, and so on.

Logogen and imagen units also differ fundamentally in their meaningfulness. Modality-specific logogen have no meaning in the semantic sense that characterizes the standard views of linguistic lexical representations. They are directly “meaningful” only in that, when activated, they have some degree of familiarity or recognizability. Imagens, however,
are intrinsically meaningful in that the conscious images they activate resemble the perceived objects and scenes they represent. Further meaning for both logogens and imagens arises from their referential or associative connections to other representations of either class. Referential connections between concrete-word logogens and imagens permit objects to be named and names to activate images that represent world knowledge. Associative connections between logogens (whether concrete or abstract) and between imagens allow for within-system associative processing that defines associative meaning as measured, for example, by word association tests and analogous nonverbal procedures. All DCT interunit connections are many-to-many and their activation is probabilistically determined by task-relevant and contextual stimuli.

The preceding statement means that contextual relations and task-relevant properties of items are described in terms of the same DCT concepts, namely, logogens, imagens, and connections between them. For example, mental images provide situational contexts for language in the absence of the referent objects and settings. Verbal associations include meaningful relationships such as synonymy, antonymy, and paraphrases as well as intralinguistic contextual relations. Ensembles of DCT units and contextual relations are involved in different tasks such as learning word pairs or understanding sentences and longer texts. The tasks might involve extraneous contexts, such as task instructions, that vary in their relations to the task-relevant ensembles so that they could enhance, interfere with, or have no effect on task performance. These implications differ from those arising from Elman’s context-dependency proposal. Empirical evidence (reviewed later) bears on the alternatives.

We conclude this section by identifying parallels to DCT assumptions in other theories. (a) Regarding logogen length, linguistic theories generally accept idioms and fixed phrases as lexical units, and Langacker (1991) suggested further that the “lexicon is most usefully...described as the set of fixed expressions in a language, irrespective of size and regularity. Thereby included as lexical items are morphemes, stems, words, compounds, phrases, and even longer expressions—provided that they are learned as established units” (cited in Schönefeld, 2001, p. 182). (b) The limited meaningfulness of logogens agrees with the assumption proposed by Rumelhart (1979) and accepted by Elman (2004) that lexical words are clues to meaning rather than being semantically meaningful in themselves. (c) The role of interword associations on the verbal side of DCT has its parallel in corpus linguistic studies of word collocations in text and speech. Psychological correspondences include Kiss’s (1975) associative thesaurus of English and the latent semantic analysis of large language corpuses by Landauer and Dumais (1997). (d) The DCT multimodal representational systems has a partial parallel in Morton’s (1979) logogen theory, including its addition of imagen-like “pictogens” to account for picture recognition. Likewise, Coltheart (2004) found neuropsychological evidence for at least two lexicons, visual and auditory, along with a nonverbal “lexicon” involved in object recognition. Similarly, Caramazza (1997) postulated modality-specific lexical forms in lexical access without any modality-neutral (lemma) level of lexical representation. (e) Finally, the DCT view of context is generally similar to that of other theories, but with specific aspects that are unique to DCT.
3. The nature of event knowledge

Elman (2009) appropriately emphasized event knowledge in relation to current language behavior. Such knowledge was crucial as well in the evolution of syntactic communication (Nowak, Plotkin, & Jansen, 2000). Syntactic properties became part of language “because they can activate useful memories of relevant events in both listener and speaker in the absence of the perceptual events themselves” (Paivio, 2007, p. 308). The main problem, then, is how the absent events are cognitively represented. First, event discourse and nonverbal event knowledge are necessarily distinct, as Elman implied when he asked about “the nature of the events being described” (Elman, 2009, p. 20). However, he then went on to treat events operationally as verbal descriptions rather than as nonverbal representations of the described events. Thus, meanings, roles, contexts, and ambiguities of target sentences entailed language properties that affect, for example, expectations of what words are likely to occur next during sentence processing. Elman assumed that such effects are mediated by representations of events involving referent objects but provided no independent defining operations for such mediators.

Rather than concretize event knowledge, Elman sought an abstract form of representation by first adding conceptual knowledge to the distinction between world knowledge and linguistic knowledge. He questioned whether the tripartite differences “require a separate copy of a person’s conceptual and world knowledge plus the linguistic facts that are specific to linguistic forms. Or is there some other way by which such (differences) can operate directly on a shared representation of conceptual and world knowledge?” (p. 22).

The “other way” is via event schemas. He asserted that “The similarity between event knowledge and schema theory should be apparent” (p. 22). However, this similarity is not apparent. What constitutes an “event” or a “schema” is not well defined. Elman suggested that schemas are epiphenomenal, emerging “as a result of (possibly higher order) co-occurrence between the various participants in the schema. The same network may instantiate multiple schemas, and different schemas may blend. Finally, schemas emerge as generalizations across multiple individual examples” (p. 24). What the “participants” are, how they network, and how this network of participants can both produce and instantiate multiple epiphenomena that both generalize and blend meaning is not apparent in any version of schema theory.

The shortcomings of schema as an explanatory concept have been identified in a number of reviews (e.g., Alba & Hasher, 1983; Sadoski, Paivio, & Goetz, 1991). These reviews concluded that schema theory fails to account for the rich and detailed memory of complex events frequently observed in research. The cognitive puzzle here is how specific objects and events are transformed into abstract representations, from which the original details are somehow recovered later at a better than chance level. This fatal empirical and logical instantiation problem does not arise in DCT because it uses theoretically and operationally defined modality-specific representations to predict and explain performance in cognitive tasks.
4. Empirical implications and evidence

Elman (2009, pp. 17–21) summarized collaborative research that tested predictions from his contextual approach using meaningful but unexpected combinations of roles played by agents, instruments, and patients in event descriptions. The results showed that the unexpected combinations (e.g., *the butcher cut the grass with scissors*) slow down on-line processing relative to expected combinations (e.g., *the butcher cut the meat with a knife*). The studies provided new information on such details as how quickly the effects occur as the sentence unfolds, and converge theoretically with DCT’s emphasis on the anticipatory functions of dual coding systems based on anticipatory imagery and/or verbal associations (e.g., Paivio, 2007, pp. 87–90; Sadoski & Paivio, 2001, chapter 4).

The DCT explanation of the Elman et al. results described above is as follows. In actual reading, larger text and pragmatic contexts plus the sentence stem *The butcher cut the…* could evoke anticipatory verbal associates (*meat, knife*) and also anticipatory images of a butcher cutting meat with a knife. In DCT, both verbal and nonverbal contexts constrain and direct further anticipations probabilistically (Sadoski & Paivio, 2001; p. 73 ff.) The addition of the nonverbal code to anticipatory processing offers explanatory advantages over a schema-based interpretation. For example, a mental image might synchronically include inferred (i.e., anticipated) information regarding the type and cut of meat, the identity of the butcher (e.g., shop owner, industrial plant worker), the setting of the action, and other information that sets the stage for still further anticipations. Schemata would necessarily operate at a level more general than this; there is no default reason why butchers would be either shop owners or industrial plant workers, for example. However, imagery reports of text events typically include such specific, elaborated details (e.g., Krasny & Sadoski, 2008; Sadoski et al., 1991; Sadoski, Goetz, Olivarez, Lee, & Roberts, 1990). For example, Sadoski et al. (1990) found that nearly 25% of imagery reports from reading contained information elaborated beyond the text or imported from memory in a manner consistent with the constraints of the text (e.g., allowable specifics of setting, action, appearance). Furthermore, nearly 43% of such elaborations and importations were included in reports that combined information from across text segments. That is, inferred information reported in imaginal form was applied to adjacent and ensuing text. At the sentence level, converging behavioral and neuropsychological evidence indicates that mental imagery occurs during sentence processing, especially for concrete sentences (e.g., Bergen, Lindsay, Matlock, & Narayanan, 2007; Holcomb, Kounios, Anderson, & West, 1999; Paivio & Begg, 1971).

Elman conceded that lexical explanations could work if one assumes a sufficiently information-rich lexicon associated with verbs (e.g., cut) that can combine with many kinds of agents and instruments, but that lexical explanations are less plausible in the case of effects of verb aspect, which do not depend on verb-specific information.

However, given the relations (albeit complex) between verb aspect and verb tense, explanations based on DCT-defined lexical representations are plausible. First, tense-inflected verbs (e.g., eaten, ate) can be independent logogen units in DCT. Second, verb past tense can be generated from either verb stems or action pictures (Woollams, Joanisse, & Patterson, 2009), implicating DCT logogens and imagens. Third, participants can generate distinct
nonverbal images to represent situational aspects of past, present, and future events (Werner & Kaplan, 1963, pp. 425–438), the last reflecting the real-world anticipatory function of imagery alluded to above.

We turn to DCT-related research concerning questions that arise from Elman’s article, focusing on (a) the distinction between world knowledge and linguistic knowledge, (b) the functional reality of abstract conceptual representations, and (c) the lexicon-context issue.

4.1. World knowledge and linguistic knowledge

The distinction between world knowledge (the DCT nonverbal system) and linguistic knowledge (the DCT verbal system) is fundamental in DCT research. The distinction is operationally defined in terms of variables that affect the probability that verbal or nonverbal systems (or both) will be used in a given task. The most relevant classes of defining variables for present purposes are stimulus attributes (e.g., pictures versus words, concrete language versus abstract language), and experimental manipulations (e.g., instructions to use imagery or verbal strategies).

Such methods have revealed separate and joint effects of verbal and nonverbal variables in numerous tasks (most recently reviewed in Paivio, 2007). Additively independent dual coding effects on memory were obtained with materials ranging from single items, to pairs, phrases, sentences, paragraphs, and longer text. For example, presenting pictures along with their names increases recall additively relative to once-presented words or pictures, as does image-plus-verbal coding of words. A similarly large memory advantage consistently occurred for concrete over abstract verbal material, which is explainable in terms of differential dual coding resulting from a higher probability of imagery activation by concrete than abstract language. Additive dual coding effects have also been obtained in comprehension tasks (e.g., Mayer, 2001; Sadoski, Goetz, & Fritz, 1993; Sadoski, Goetz, & Rodriguez, 2000).

Singularly important here are effects predicted from the conceptual peg hypothesis of DCT (its history and current status are reviewed in Paivio, 2007, pp. 22–24, 60–67). The hypothesis states that item concreteness is especially potent when it is a property of the item that serves as the retrieval cue in associative memory. Predictions were confirmed in memory experiments which showed that concreteness of the retrieval cue was related to response recall much more strongly than concreteness of the response items, or concreteness of items in non-cued (free) recall. The relevance here is two-fold. First, the hypothesis agrees with Elman’s general emphasis on the importance of words as cues. Second, the efficacy of concrete cues is linked to their capacity to activate images of referent objects or situations that mediate recall, thus requiring direct access to knowledge of the world. Elman’s words-as-cues approach, however, does not specify mechanisms with similar predictive implications.

The conclusion is that DCT-related research reveals contributions of both nonverbal world knowledge and linguistic knowledge to language phenomena that are not revealed by Elman’s (2009) reliance on event-descriptive language materials alone. World knowledge likely played a role in his experimental results, but the extent of that contribution is uncertain because the inferred real-world relations are confounded with verbal associative
relations between words that describe agents, instruments, patients, and verbs. Even two experiments that used pictures and imagery instructions made no comparisons with analogous verbal procedures to test for differential contributions of the pictures (or imagery) and language.

Research under the rubric of embodied cognition supports the same conclusion. The prototypical studies relate language comprehension and memory to nonverbal motor processes, perception, imagery, and language. Early research summarized by Werner and Kaplan (1963, pp. 26–29) showed that, to be perceived at apparent eye level, the printed words ‘‘climbing’’ and ‘‘raising’’ had to be positioned below ‘‘lowering’’ and ‘‘dropping.’’ Werner and Kaplan interpreted the effect in terms of an organismic theory according to which word meaning exerts a directional ‘‘pull’’ consistent with the dynamic meaning of the stimulus. Many variants of the Werner et al. studies have been recently reported (e.g., Bergen et al., 2007; Glenberg & Kaschak, 2002; Zwaan, 2004). An experiment by Louwerse (2008) is especially apropos because it distinguished between effects attributable to world knowledge and to language. ‘‘Iconic’’ word pairs (attic-basement) were judged to be related faster than reverse iconic pairs when the words were presented in a vertical spatial arrangement but not when they were presented horizontally. The initial interpretation was that the effects resulted from nonverbal world knowledge of spatial relations. However, measures of relational frequency showed that iconic word order is more frequent than non-iconic word order and that, when word-order frequency was controlled, the iconicity effect disappeared. Such confounding by linguistic associations was not investigated in the event knowledge experiments by Elman and his collaborators. DCT explains these effects readily.

4.2. The functional reality of abstract representational concepts

Elman proposed that world knowledge and linguistic knowledge draw on abstract conceptual knowledge that he interpreted in terms of an improved (though as yet unrealized) version of schema theory. We have discussed schema theory critically and here we deal similarly with a broader class of abstract conceptual representations that DCT research has addressed. The relevant studies systematically varied item attributes and task characteristics designed to test effects of modality-specific processes that could not be explained by undifferentiated properties of any single, modality-neutral representational code. An early review (Paivio, 1983) turned up 60 independent findings by various researchers that were predicted or explained by DCT but not single-code theories. A more comprehensive summary (Paivio, 1986) prompted a reviewer to conclude that ‘‘The data demand something better than common coding theories have been able to provide’’ (Lockhart, 1987, p. 389). That conclusion has been further strengthened by recent findings from behavioral and neuroscience research (reviewed in Paivio, 2007).

4.3. Item-specific variables versus context

Many early DCT studies (e.g., see Paivio, 1971, pp. 377–384) explicitly investigated the joint effects of item-specific variables (e.g., pictures versus words) and contextual variables
(e.g., conjunctive versus meaningful relational organization of units). Robust item-specific memory effects were augmented by meaningful contexts. Subsequently, beginning in the 1980s, context became specifically relevant to DCT because some theorists suggested that language concreteness effects depend on contextual support that is generally more available for concrete than abstract items (e.g., Schwanenflugel & Shoben, 1983). The context-availability hypothesis is a specific variant of Elman’s more general hypothesis that lexical knowledge is context dependent. However, neither hypothesis explains the persistent concreteness effects in a wide variety of contexts in the early studies, nor in studies that controlled for context or were designed to pit contextual variables against item concreteness.

Large concreteness effects were found in comprehension and recall of sentences and paragraphs matched for verbal contextual factors (Sadoski et al., 1993, 2000). Additively independent memory effects of item concreteness/imagery and contextual variables were obtained using: (a) noun pairs (Paivio, Walsh, & Bons, 1994); (b) adjective–noun pairs and sentences (Paivio, Khan, & Begg, 2000); and (c) concrete and abstract words presented in the context of meaningful sentences or in anomalous ones that inhibited relational processing (Richardson, 2003). No hint of an interaction occurred in any of these experiments.

Sadoski, Goetz, and Avila (1995) tested competing predictions using two sets of paragraphs about historical figures and events that were matched for number of sentences, words, syllables, sentence length, information density, cohesion, and rated comprehensibility. One set of paragraphs were rated equal in familiarity but unequal in concreteness. Here, DCT predicted that the concrete paragraphs would be recalled better than the abstract paragraphs due to the advantage provided by imagery, whereas context availability theory predicted comparable recall for the two types because they were alike in familiarity and contextual support. In another set, the paragraphs differed in both familiarity and concreteness, with the abstract paragraph being relatively more familiar. In this set, DCT predicted that recall of the familiar abstract paragraph would approximate recall of the unfamiliar concrete paragraph (i.e., offsetting disadvantages), whereas context availability theory predicted that the abstract paragraph would be recalled better than the concrete paragraph (reflecting the advantage of greater familiarity). The results matched the predictions of DCT but not context availability theory.

Begg and Clark (1975) obtained imagery ratings for homonyms that have both a concrete and an abstract meaning, as well for the words in sentence contexts that biased concrete or abstract interpretations (e.g., justice of the peace versus love of justice). Free recall tests showed that out-of-context word imagery ratings correlated significantly with recall of words in lists, whereas imagery ratings in contexts correlated significantly with recall of the words in sentence contexts. Thus, the experiment demonstrated both item-specific and context-dependent effects of imageability.

O’Neill and Paivio (1978) showed interactive effects of concreteness and extreme variation of context on comprehension, imagery, and memory. Ratings of comprehensibility, sensibleness, and imagery were obtained for normal concrete and abstract sentences as well as anomalous sentences created by substituting content words from one sentence to another. The substitutions produced general rating decrements on all variables, but the decrements were greater for concrete than abstract sentences. Most notably, whereas comprehensibility
and sensibleness ratings were higher for concrete than abstract normal sentences, the difference was completely reversed for anomalous sentences. Moreover, an incidental free recall task following the ratings showed that recall of content words and whole sentences was much higher for concrete than abstract materials whether sensible or anomalous, and word imageability specifically benefited recall in anomalous as well as meaningful contexts, presumably because the words evoked memorable images in either case. Thus, DCT item-specific variables benefited recall even in massively disrupted contexts.

In sum, the studies cited in this section revealed persistent effects of DCT-relevant item-specific lexical variables that were sometimes qualified by contextual variables in ways predictable from DCT. The results appear not to be explainable in terms of Elman’s suggestion that behavioral effects of lexical knowledge arise mainly from language contexts in which lexical units occur.

5. Computational modeling

Elman’s (2009) stated goal is to develop a computational model that would be consistent with a contextual explanation of apparent lexical influences on sentence processing. He conceded that his Simple Recurrent Network model is too simple to serve as anything but a conceptual metaphor and he envisages modeling event schemas using newer connectionist architecture that is ‘‘better suited for instantiating the schema’’ (p. 23). Thus far this goal remains a promissory note and in our view it is likely to remain elusive because it requires modeling an abstract conceptual entity that has not been successfully instantiated in terms of empirical correlates. The theory-building enterprise moves from observable sentence phenomena to assumed knowledge of the world to increasingly abstract descriptions and conceptual representations, ending with completely disembodied computational models. We await the development and domain-specific explanatory power of the newer improved models.

We turn finally to DCT and computational modeling. A useful bridge to the topic is the situational model in Kintsch’s theory of comprehension because, like Elman and DCT, the situational model is intended to represent knowledge of the world, including event sequences. Moreover, as in Elman but not DCT, the situation model is represented in an abstract, propositional format related to schemas. It is therefore somewhat surprising to see his recent concession that ‘‘Situational models may be imagery based, in which case the propositional formalism used by most models fails us’’ (Kintsch, 2004, p. 1284; see also Kintsch, 1998).

A computational escape from the above impasse would require direct formal modeling of nonverbal event knowledge as reflected in imagery and pictured scenes. The models to date have failed to represent the kinds of detailed event information that behavioral experiments have shown to be available perceptually and in language-evoked imagery. The impasse is the same as that involved in Elman’s (2009) and Kintsch’s (2004) representation of event knowledge only indirectly as event descriptions. That is, computational scene perception and imagery are models based on natural-language verbal descriptions transformed into
abstract formal descriptions (e.g., propositions, structural descriptions) that are necessary for computer programming. This was the case with early imagery simulations and it remains so in more recent computational models of static and dynamic imagery (e.g., Croft & Thagard, 2002) as well as AI-inspired computational imagination (Setchi, Lagos, & Froud, 2007). Problems associated with computer simulation models motivated Kosslyn to abandon the computer metaphor and shift instead to tests of a theory of imagery based on functional properties of the brain (Kosslyn, Van Kleek, & Kirby, 1990).

A possible exception to this negative conclusion is Mel’s (1986) use of virtual robotics together with connectionist architecture to model three-dimensional mental rotation, zoom, and pan. Using a flat array of processors driven by a coarsely tuned binocular feature detector, the system learned to run simulations of the visual transformations from visual-motor experience with various kinds of motion. It remains to be seen how far the approach can be extended to comprehension, memory, and other phenomena relevant to DCT or Elman’s approach, and also go beyond simulation of known effects to generate predictions and discoveries of new properties of imagery, perception, and language.

6. Conclusions

We conclude that there is much value in Elman’s reconceptualization of the mental lexicon and in his emphasis on contextualized event knowledge. However, we do not agree that schema theory or other models based on computational descriptions offer an adequate solution to the issues he raises. The main problem is that such theories do not include the multimodal, verbal–nonverbal distinctions necessary for capturing the richness of real-world contexts that we agree are needed to fully account for meaning. Only theories that deal directly with these distinctions would be sufficient, and we submit DCT as one viable candidate (Sadoski & Paivio, 2007).

References


