Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 53 (2005) 258–276
Journal of Memory and Language
www.elsevier.com/locate/jml

A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning

Steve R. Howell *, Damian Jankowicz, Suzanna Becker

McMaster University, Hamilton, Ont., Canada

Received 1 December 2004; revision received 15 March 2005
Available online 25 April 2005

Abstract

It is generally accepted that children have sensorimotor mental representations for concepts even before they learn the words for those concepts. We argue that these prelinguistic and embodied concepts direct and ground word learning, such that early concepts provide scaffolding by which later word learning, and even grammar learning, is enabled and facilitated. We gathered numerical ratings of the sensorimotor features of many early words (352 nouns, 90 verbs) using adult human participants. We analyzed the ratings to demonstrate their ability to capture the embodied *meaning* of the underlying concepts. Then using a simulation experiment we demonstrated that with language corpora of sufficient complexity, neural network (SRN) models with sensorimotor features perform significantly better than models without features, as evidenced by their ability to perform word prediction, an aspect of grammar. We also discuss the possibility of indirect acquisition of grounded meaning through "propagation of grounding" for novel words in these networks.

Keywords: Language acquisition; Features; Semantics; SRN; Neural network; Sensorimotor; Conceptual learning

Considerable evidence suggests that by the time children first begin to learn words around the age of 10–12 months, they have already acquired a fair amount of sensorimotor (sensory/perceptual and motor/physical) knowledge about the environment (e.g., Lakoff, 1987; Lakoff & Johnson, 1999; Bloom, 2000; Langer, 2001), especially about objects and their physical and perceptual properties. By this age children are generally able to manipulate objects, navigate around their environment, and attend to salient features of the world, including parental gaze and other cues important for word learning (Bloom, 2000). Some have suggested that this pre-linguistic conceptual knowledge has a considerable effect on the processes of language acquisition (Lakoff, 1987; Mandler, 1992; Smith & Jones, 1993) and even on later language processing (e.g., Glenberg & Kaschak, 2002; Barsalou, 1999). We also argue that the evidence indicates that this early prelinguistic knowledge has great impact, directly and indirectly, throughout a number of phases of language learning, and we attempt to begin to demonstrate this with a neural network model.

* Corresponding author. Present address: Department of Psychology (WJ Brogden Hall), University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53703, USA. Fax: +1 608 262 4029.
E-mail address: showell@wisc.edu (S.R. Howell).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2005.03.002

To begin with, this prelinguistic conceptual information helps children to learn their first words, which correspond to the most salient and imageable (Gillette, Gleitman, Gleitman, & Lederer, 1999) objects and actions in their environment, the ones they have the most experience with physically and perceptually. Generally speaking, the more "concrete" or "imageable" a word, the earlier it will be learned. This helps to explain the preponderance of nouns in children's early vocabularies (see Gentner, 1982). The meanings of verbs are simply more difficult to infer from context, as discussed by as demonstrated by Gillette et al. (1999). Only the most clearly observable or "concrete" verbs make it into children's early vocabularies. However, later verbs are acquired through the assistance of earlier-learned nouns. If a language learner hears a simple sentence describing a real-world situation, such as a dog chasing a cat, and already knows the words *dog* and *cat*, the only remaining word must be describing the event, especially if the learner already has built up a pre-linguistic concept of "dogs chasing cats" at the purely observational level. As Bloom (2000) describes, the best evidence for "fast-mapping" or one-shot learning of words in children comes from similar situations in which only one word in an utterance is unknown, and it has a clear, previously unknown, physical referent present. Of course, since the verb *chase* refers to an event rather than an object, the above example is not an exact fit to the fast-mapping phenomenon as it is usually described, but it is similar.

These very first words that children learn thus help constrain the under-determined associations between the words children hear and the objects and events in their environment, and help children to successfully map new words to their proper referents. This happens through the use of cognitive heuristics such as the idea that a given object has one and only one name (Markman & Wachtel, 1988), or more basic object-concept primitives (Bloom, 2000) such as object constancy.With a critical mass of some 50 words, children begin to learn *how to learn* new words, using heuristics such as the count-noun frame, or the adjective frame (Smith, 1999). These frames are consistent sentence formats often used by care-givers that enable accurate inference on the part of the child as to the meaning of the framed word, e.g., "This is a ___." These factors combine to produce a large increase in children's lexical learning at around 20 months. As they begin to reach another critical mass of words in their lexicon (approaching 300 words), they start to put words together with other words—the beginnings of expressive grammar (Bates & Goodman, 1999). Around 28 months of age children enter a "grammar burst" in which they rapidly acquire more knowledge of the syntax and grammar of their language, and continue to approach mature performance over the next few years.

By this account of language acquisition, conceptual development has primacy; it sets the foundation for the language learning that will follow. Words are given meaning quite simply, by their associations to real-world, perceivable events. Words are directly *grounded* in embodied meaning, at least for the earliest words. Of course, it may not be just simple statistical associations between concepts and words in the environment; the child is an active learner, and processes like joint attention or theory of mind may greatly facilitate the learning of word to meaning mappings (Bloom, 2000).

Of course, it seems clear that the incredible word-learning rates displayed by older children (Bloom, 2000) indicate that words are also acquired by linguistic context, through their relations to other words. Children simply are learning so many new words each day that it seems impossible that they are being exposed to the referents of each new word directly. The meanings of these later words, and most of the more abstract, less imageable words we learn as adults, must clearly be acquired primarily by their relationships to other known words. It may in fact be true that these meanings can *only* be acquired indirectly, through relationships established to the meanings of other words.

Evidence for the indirect acquisition of meaning is not limited to the speed with which children learn words. The work of Landauer and colleagues (e.g., Landauer and Dumais, 1997; Landauer, Laham, & Foltz, 1998) provides perhaps the clearest demonstration that word "meanings" can be learned solely from word-to-word relationships (although see Burgess & Lund, 2000; for a different method called HAL). Landauer's Latent Semantic Analysis (LSA) technique takes a large corpus of text, such as a book or encyclopedia, and creates a matrix of co-occurrence statistics for words in relation to the paragraphs in which they occur. Applying singular-value decomposition to this matrix allows one to map the words into a high-dimensional space with dimensions ordered by significance. This high-dimensional representation is then reduced to a more manageable number of dimensions, usually 300 or so, by discarding the least significant dimensions. The resulting compressed meaning vectors have been used by Landauer et. al. in many human language tasks, such as multiple choice vocabulary tests, domain knowledge tests, or grading of student exams. In all these cases, the LSA representations demonstrated human-level performance.

While models based on these high-dimensional representations of meaning such as LSA and HAL perform well on real world tasks, using realistically sized vocabularies and natural human training corpora, they do have several drawbacks. First, they lack any consideration of syntax, since the words are treated as unordered collections (a 'bag of words'). Second, LSA and HAL meaning vectors lack any of the grounding in reality that comes naturally to a human language learner. Experi-