Cognitive Psychology 97 (2017) 1–30

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

An associative account of the development of word learning

Vladimir M. Sloutsky [a,*], Hyungwook Yim [a], Xin Yao [b], Simon Dennis [c]

[a] Department of Psychology, The Ohio State University, United States
[b] Unilever Corporation, Shanghai, China
[c] School of Psychology, University of Newcastle, Australia

ARTICLE INFO

Article history:
Accepted 2 June 2017
Available online 20 June 2017

Keywords:
Word learning
Cognitive development
Semantic development

ABSTRACT

Word learning is a notoriously difficult induction problem because meaning is underdetermined by positive examples. How do children solve this problem? Some have argued that word learning is achieved by means of inference: young word learners rely on a number of assumptions that reduce the overall hypothesis space by favoring some meanings over others. However, these approaches have difficulty explaining how words are learned from conversations or text, without pointing or explicit instruction. In this research, we propose an associative mechanism that can account for such learning. In a series of experiments, 4-year-olds and adults were presented with sets of words that included a single nonsense word (e.g. dax). Some lists were taxonomic (i.,e., all items were members of a given category), some were associative (i.e., all items were associates of a given category, but not members), and some were mixed. Participants were asked to indicate whether the nonsense word was an animal or an artifact. Adults exhibited evidence of learning when lists consisted of either associatively or taxonomically related items. In contrast, children exhibited evidence of word learning only when lists consisted of associatively related items. These results present challenges to several extant models of word learning, and a new model based on the distinction between syntagmatic and paradigmatic associations is proposed.

1. Introduction

The ability to learn and use words is a fundamental aspect of the human cognitive system. Not only is word learning virtually universal among humans, but it also exhibits an early onset: children typically reach an average vocabulary of 200–300 words by 24 months (Fenson et al., 1994). Despite its early onset, from the logical point of view, word learning should be very difficult, as the learner has to solve a number of hard problems. First, when presented with a novel word accompanying a novel item, the young word learner has to solve a *mapping* problem – how to map the word in question onto the world. For example, as noted by Quine (1960), the word *gavagai* uttered while pointing to a rabbit has an infinitely large number of potential mappings, including the rabbit, its parts, its texture, its method of locomotion, or the speaker's attitude towards the rabbit, among many others. Second, the word learner has to solve a *generalization* problem – given that most of the time words can be extended to multiple tokens, it is necessary to determine the class of entities the word refers to. For example, when an object (say a rabbit) is named, even if the mapping problem is solved, the word *gavagai* has an large number of potential extensions: this particular rabbit, some rabbits, all rabbits, rabbits and cats, land animals, mammals,

* Corresponding author at: Department of Psychology, The Ohio State University, Columbus, OH 43210, United States.
E-mail address: sloutsky.1@osu.edu (V.M. Sloutsky).

http://dx.doi.org/10.1016/j.cogpsych.2017.06.001
0010-0285/

all animals, solid objects, all things, etc. Therefore, both mapping and generalization are massively underdetermined by input, even when a to-be-named object is present, pointed to, and explicitly named (such situations are often referred to as ostensive definitions).

The situation appears to be even more difficult and uncertain when a word is introduced only in a conversation or through reading (as we argue, most words are, see Gleitman, 1990, for related arguments and evidence), without pointing or explicit naming. This uncertainty persists even if we limit our focus (which we do in present work) to word forms that are the earliest to be acquired – count nouns. Despite this massive uncertainty, most children succeed in learning words. Therefore, it is reasonable to ask: How do they accomplish this?

The goal of the work reported here is to provide answers to this question and to consider our answers in the context of previously provided ones. Given the history and the importance of the problem, it is hardly surprising that there are multiple accounts of word learning (see Regier, 2005, for an excellent review). In what follows, we briefly review two classes of accounts, inference-based and association-based. We also discuss why some of these accounts may have difficulty explaining an important aspect of word learning -- learning words from context, including conversations or reading. We then present our proposal, a set of experiments designed to test ours as well as some of the previously proposed accounts, and a computational model instantiating this proposal.

1.1. Inference-based models

The first set of approaches considers word learning as an inference problem: the learner needs to select the most appropriate hypothesis as to what the word in question might mean (see Bloom, 2000; Xu & Tenenbaum, 2007, for reviews). The basic idea is that despite massive uncertainty (i.e., the same data set is compatible with multiple hypotheses), the problem can be solved if (a) one is biased in favor of some hypotheses over others and/or (b) hypotheses are weighted in terms of their support by data. An important commonality of inference-based models is that determining the meaning of a word is driven by a set of assumptions: these assumptions help the learner to converge on the correct meaning.

One proposal is that even very young word learners use their knowledge about the language and the world to select a correct word-world mapping (Markman, 1989; Xu & Tenenbaum, 2007). Some proponents of this idea suggest that word learners hold certain assumptions that constrain the number of possible word-world mappings. In particular, young children are believed to assume that count nouns (a) denote whole objects rather than parts and (b) refer to taxonomic kinds rather than to thematic groupings (Markman, 1990; Markman & Hutchinson, 1984). Therefore, if a novel object is labeled (say a dog is labeled by a word *dog*) a young word learner would assume that the word refers to the whole animal, and not to its parts. Furthermore, when presented with a novel dog, the learner would apply the assumption that words refer to taxonomic kinds (rather than to individual objects or to thematic groupings), thus extending the word *dog* to a novel instance. These assumptions should substantially reduce the hypothesis space, thereby facilitating word learning.

Although these assumptions could be helpful for solving the mapping problem, the idea of young word learners having these assumptions (or constraints) has several substantial limitations (e.g., see McMurray, Horst, & Samuelson, 2012, for a recent set of critiques). Where do these assumptions come from? How do these assumptions get relaxed when children learn synonyms, words for superordinate categories, or adjectives? And how do conflicts among the assumptions get resolved?

In addition, while these constraints are useful for solving the mapping problem, their ability to solve the generalization problem is less clear (see Xu & Tenenbaum, 2007, for a discussion). In particular, while the taxonomic assumption may suggest to the learner that the word *dog* refers to a category (and not to an individual or to a thematic grouping), the assumption offers little help in determining where the boundaries of this category are.

A proposal by Xu and Tenenbaum (2007) has complementary strengths – it offers a solution to the generalization problem by proposing a way of weighing and comparing hypotheses. According to this proposal, young word learners solve the generalization problem by using labeling data in conjunction with "sampling" assumptions (i.e., assumptions of how the teacher samples examples during naming) as well as assumptions about the structure of categories in the world. These assumptions determine prior probabilities of what a given word might mean. In particular, young word learners assume that entities belong to categories and that categories form taxonomic hierarchies (i.e., each item can be categorized at the subordinate, basic, and superordinate levels). When an object is labeled with a novel word, the learner has to decide among the possible hypotheses, and knowledge of the conceptual hierarchy determines the prior probability of each hypothesis under consideration.

These decisions are based on a Bayesian computation -- weighing each hypothesis by its prior and its likelihood. The prior [P(h)] reflects the learner's expectations about possible word meanings independently of the data. The likelihood [P(X|h)] reflects which data patterns are likely to be observed under each particular hypothesis. The posterior probability of each hypothesis [P(h|X)] reflects the learner's subsequent belief about the word meaning and is calculated by multiplying the prior by the likelihood.

As argued by Xu and Tenenbaum (2007), priors favor more distinctive and natural hypotheses (e.g., *cats* vs. *things with black spots*), whereas likelihoods favor hypotheses that most closely correspond to the data. As a result, under typical conditions, the distinctive and narrow classes should be favored over broader and/or less distinctive classes. For example, if a single Labrador is labeled by the word *dax*, a child would decide that the word *dax* refers to Labradors and, possibly, to other dogs: while the likelihoods may somewhat favoring the Labradors, priors favor the dogs (although both are distinctive classes, the latter could be construed as more distinctive as it is the basic-level category). However, if three Labradors are