ARTICLE IN PRESS

New Ideas in Psychology xxx (2017) 1–7

Contents lists available at ScienceDirect

New Ideas in Psychology

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

A communicative approach to early word learning

Daniel Yurovsky

Department of Psychology, University of Chicago, 5848 S University Ave., Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Received 16 February 2017
Received in revised form
31 August 2017
Accepted 1 September 2017
Available online xxx

Keywords:
Language acquisition
Learning
Cognitive development

ABSTRACT

Young children learn the meanings of thousands of words by the time they can run down the street. Many efforts to explain this rapid development begin by assuming that the computational-level problem being solved is acquisition. Consequently, work in this line has sought to understand how children infer the meanings of words from cues in the communicative signals of the speakers around them. I will argue, however, that this formulation of the problem is backwards: the computational problem is communication, and language acquisition provides cues about how to communicate successfully. Under this framing, the natural unit of analysis is not the child, but the parent-child dyad. A necessary consequence of this shift is the realization that the statistical structure of the input to the child is itself dependent on the child. This dependency radically simplifies the computational problem of learning and using language.

> The infant's Language Acquisition Device could not function without the aid given by an adult who enters with him into a transactional format. That format, initially under the control of the adult, provides a Language Acquisition Support System, LASS. It frames or structures the input of language and interaction to the child's Language Acquisition Device in a manner to “make the system function.”
>
> Bruner (1983).

1. Introduction

Humans get a lot of language learning done in strikingly little time. A useful comparison here is the relative rate of two of the most chronicled developmental milestones: language and locomotion. By the time she is a year old, the descriptive norm for a typically developing infant is to produce several words, and to know the names of many common objects. However, the same infant will still struggle to walk at all unless she is holding onto furniture with one hand. When this descriptively normative infant develops into a three-year-old toddler, she will be expected to produce multi-word utterances, to understand prepositions (e.g. on, under), and to describe scenes in picture books. However, this same toddler will still be unable to stand on one foot for more than one second at a time (Squires, Bricker, & Twombly, 2009). There is every reason to think that learning to walk should be a hard problem—it certainly has been difficult to build artificial systems that do it well (e.g Collins, Ruina, Tedrake, & Wisse, 2005). Walking is a problem that humans do not seem especially adept at solving relative to other species, particularly in comparison to their clearly unique trajectory in acquiring language (Capaday, 2002; Garwicz, Christensson, & Psouni, 2009; Hockett, 1959). In contrast, human children are uniquely adept at acquiring natural language—a hard problem that infants make look easy. Indeed, in the foreword to his seminal book on the topic, Bloom (2000) writes that “the child's ability to learn new words is nothing short of miraculous.”

So what explains our precocious ability to acquire language? For the present paper, let us follow Bloom (2000) and focus specifically on learning words. And let us get even more specific: Concrete nouns. Of course, this does not exhaust the space of what children can or do learn in their first few years. But concrete nouns are a useful locus for two reasons: (1) Concrete nouns do make up a large slice of early vocabularies (Caselli et al., 1995; Gentner, 1982), and (2) The problem of acquisition should be even worse for more complex and abstract units of language.

2. The computational problem of language learning

Although details vary from analysis to analysis, roughly speaking there is broad consensus about the “computational problem of word learning” for concrete nouns (Marr, 1982). The child is an observer in a world filled with three kinds of

E-mail address: yurovsky@uchicago.edu.

http://dx.doi.org/10.1016/j.newideapsych.2017.09.001
0732-118X/

observables: words, objects, and referential cues. On any given occasion, the child hears a subset of the words, sees a subset of the objects, and also possibly sees one or more referential cues (e.g. a speaker's gaze) that point to a subset of the objects. The computational problem is to recover from these observables a lexicon—a latent structure that details the mapping between words and objects. The solution to this problem is to resolve the uncertainty about the lexicon by leveraging either the cues available on individual instances, the statistical relationship between words and referents across instances, or both (e.g., Blythe, Smith, & Smith, 2010; Frank, Tenenbaum, & Fernald, 2013; Kachergis, Yu, & Shiffrin, 2012; McMurray, Horst, & Samuelson, 2012; Siskind, 1996; Yu, 2008; Yurovsky, Fricker, Yu, & Smith, 2014; etc.).

Following this analysis, there is a growing body of experimental evidence that humans—both adults and children—are capable of using exactly this kind of information to learn words. For instance, infants are sensitive to cues like eye-gaze and pointing quite early in life, and can be shown reliably to use them to learn novel words early in the second year of life (e.g. Baldwin, 1993; Corkum & Moore, 1998; Scaife & Bruner, 1975; Tomasello, Carpenter, & Liszkowski, 2007). Similarly, adults have been shown to infer word-object mappings from co-occurrence information under a host of different conditions (e.g. Vouloumanos, 2008; Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013b), and many of these experiments have been extended to children and infants as well (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Vouloumanos & Werker, 2009).

Taken together, these and other similar results are taken as compelling evidence of movement in the right direction: Towards modeling the rapid pace of children's early word learning. There are skeptical arguments about this framework from the perspective of generalizability—e.g., will these same kinds of mechanisms explain the acquisition of verbs or adjectives (c.f. Scott & Fischer, 2012)? In this article, I will make a different kind of argument: Our optimism is misguided because of an unlicensed inference from early competence to expert performance (Chomsky, 1965). These and other demonstrations of early success in learning words from social and statistical cues are evidence of competence: they show that infants *can* learn from these regularities. But they have also been taken as evidence that humans excel at learning from these kinds of regularities—that they are subject to few performance constraints—and this inference is unwarranted. Many of these studies demonstrate that adults are not terribly good at learning words from social or statistical cues. And children are even worse.

Let us consider a representative case of social cues: The use of a speaker's eye-gaze to determine the target of her reference. As the title of their landmark paper says, Scaife and Bruner (1975) demonstrate the "capacity for joint visual attention in the infant." Their results show, for instance, that 30% of 2–4 month old children follow an experimenter's gaze in one or both trials on which they are tested; infants do not show levels of success near 100% until they are a year old. These studies demonstrate capacity; they do not demonstrate excellence. More recent studies using different paradigms show similar results: Young children succeed at above-chance levels, but there is significant development well into late childhood (Hollich et al., 2000; Moore, Angelopoulos, & Bennett, 1999; Yurovsky & Frank, 2017; Yurovsky, Wade, & Frank, 2013a). In all of these paradigms, success is defined as the ability to use the speaker's gaze and head direction to distinguish whether she is referring to an object on her left or an object on her right. In more complex visual settings, even older children and adults have difficulty using gaze to infer the target of a speaker's reference (Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Vida & Maurer, 2012).

The pattern of results for statistical word learning is strikingly similar. While infants demonstrate sensitivity to the co-occurrence information between words and objects, their memory for this information is quite fragile, even under low levels of ambiguity. For instance, in a study by Vlach and Johnson (2013), 16-month-old infants were able to learn word-object mappings through co-occurrence statistics only when the multiple occurrences of each word were blocked, but not when exposures to different words were interleaved. Vouloumanos, Martin, and Onishi (2014) found that 18-month-olds could distinguish words that had co-occurred many times with an object from those that had never co-occurred with that object, but could not distinguish words that had co-occurred 8 times with an object from those that had co-occurred twice with it (in contrast to adults, Vouloumanos, 2008). Even for adults, however, this process of statistical inference appears to be highly constrained by limits on memory and attention (Smith, Smith, & Blythe, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Yurovsky & Frank, 2015). As the number of referents available increases, adults track less and less information about each, and their ability to recall this information falls off precipitously with time between exposure and test. In contrast to domains like low-level vision, where human performance is often quite well described by ideal observer models (e.g., Najemnik & Geisler, 2005), human statistical word learning is markedly less efficient than should be expected from a system that makes optimal use of the available information (Frank, Goodman, & Tenenbaum, 2009; Yu & Smith, 2012b; Yurovsky & Frank, 2015). Several recent papers have shown that, under some working assumptions, human-scale lexicons are learnable from statistical dependencies between words and objects from approximately the amount of words heard by typically developing children (Blythe et al., 2010, 2016). However, there is little in the way of guarantees in these models that learning will be rapid (Reisenauer, Smith, & Blythe, 2013; Vogt, 2012), especially under the kinds of memory and attentional constraints found in young infants.

One should not conclude from this data that social cues and statistical cues are not useful for word learning, nor should one conclude that children do not use social cues or do not use statistical information to learn relationships between the words of their native language and the objects in the world. But the discrepancy between children's competence under ideal circumstances and their performance under more challenging circumstances raises a question: Why do children learn words so rapidly even though their learning performance is so constrained? The solution, I will argue, is that our consensus about the computational problem of word learning is incorrect. The right question is not "how do children learn the meanings of words," but rather "how do children and their parents develop systems for communicating successfully?" (Bruner, 1975). Put another way, we often think of the lexicon as the goal and the communicative moments as the tools through which the lexicon is acquired. I propose that we should make progress instead by inverting this relationship: Communication is the goal, and the lexicon is a tool for successful communication.

3. The computational problem of communication

The computational level description of word learning implicitly makes a strange kind of division: It divorces the problem of learning words entirely from the problem of using them; it assumes that the lexicon is a static property of the external world. That is, that there is some objectively "right" mapping between a word and its meaning in the same way that there is a "right" way to walk (c.f. Tomasello, 2001). But these are two very different kinds of problems. The solution for the problem of walking is constrained by biomechanics—the best way to walk is one that minimizes energy expenditure and probability of falling while maximizing distance traversed per unit time. Further, the right way to walk does not

Please cite this article in press as: Yurovsky, D., A communicative approach to early word learning, New Ideas in Psychology (2017), http://dx.doi.org/10.1016/j.newideapsych.2017.09.001