Cognition 170 (2018) 64–75

Contents lists available at ScienceDirect

Cognition

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

Original Articles

The role of reference in cross-situational word learning☆

Felix Hao Wang[a,*], Toben H. Mintz[a,b]

[a] Department of Psychology, University of Southern California, United States
[b] Department of Linguistics, University of Southern California, United States

ARTICLE INFO

Keywords:
Language acquisition
Word learning
Statistical learning
Hypothesis testing

ABSTRACT

Word learning involves massive ambiguity, since in a particular encounter with a novel word, there are an unlimited number of potential referents. One proposal for how learners surmount the problem of ambiguity is that learners use cross-situational statistics to constrain the ambiguity: When a word and its referent co-occur across multiple situations, learners will associate the word with the correct referent. Yu and Smith (2007) propose that these co-occurrence statistics are sufficient for word-to-referent mapping. Alternative accounts hold that co-occurrence statistics alone are insufficient to support learning, and that learners are further guided by knowledge that words are referential (e.g., Waxman & Gelman, 2009). However, no behavioral word learning studies we are aware of explicitly manipulate subjects' prior assumptions about the role of the words in the experiments in order to test the influence of these assumptions. In this study, we directly test whether, when faced with referential ambiguity, co-occurrence statistics are sufficient for word-to-referent mappings in adult word-learners. Across a series of cross-situational learning experiments, we varied the degree to which there was support for the notion that the words were referential. At the same time, the statistical information about the words' meanings was held constant. When we overrode support for the notion that words were referential, subjects failed to learn the word-to-referent mappings, but otherwise they succeeded. Thus, cross-situational statistics were useful only when learners had the goal of discovering mappings between words and referents. We discuss the implications of these results for theories of word learning in children's language acquisition.

1. Introduction

How language learners acquire the meanings of words is of central importance to understanding the initial stages of language acquisition, as words are the fundamental meaning-bearing units of language, and the raw materials that grammars manipulate. Moreover, decades of research have shown that lexical and grammatical acquisition are fundamentally intertwined (Fisher, Gleitman, & Gleitman, 1991; Gleitman, 1990; Pinker, 1989), even though the nature of this complex relationship is a topic of debate. Advances in understanding the mechanisms of word learning thus contribute to our understanding of how fundamental elements of language are acquired, which in turn offers additional constrains on theories of grammatical acquisition.

Word learning is also of theoretical interest because it poses an extremely challenging computational problem for the child, due to *referential ambiguity*. Ambiguity arises because any time a word is uttered, there is a limitless array of objects, relations, and states that speakers could have in mind (Gleitman, 1990; Medina, Snedeker, Trueswell, & Gleitman, 2011; Quine, 1960). Ambiguity is present even under the simplifying assumption that the intended referent is available perceptually when the learner hears a novel word, ignoring cases of reference to absent entities, or abstract concepts. The sources of referential ambiguity can be categorized into at least two distinct types. The first type of ambiguity, and the one we take up here, involves identifying the physical manifestation of the referent in the environment (i.e., the object, relation or state) that was intended as the word's referent. The other degree of freedom (or ambiguity) arises from how an entity—object, state, or event—should be linked to conceptual representations. For example, once a learner determines that a word refers to a particular object, such as a toy mouse, there is still the problem of determining which of the many possible conceptual representations of that object, the speaker has in mind: MOUSE, MAMMAL, ANIMAL, TOY, WOODEN-OBJECT, CAT-PREY, FURY-THING, PATIENT-OF-ACTION, etc. The computational challenge is in narrowing down the vast array of referents and concepts to just the right one—something at which children are very adept (Carey & Bartlett, 1978; Markman, 1990). Understanding how learners surmount this challenge is informative about mechanisms in word learning, but also could have

☆ Portions of the work were presented at the 28th Annual CUNY Conference on Human Sentence Processing. We thank John Trueswell for the many helpful discussions pertaining to this paper.

* Corresponding author at: Department of Psychology, University of Pennsylvania, 425 S University Ave, Philadelphia, PA 19104, United States.
E-mail address: haowang1@sas.upenn.edu (F.H. Wang).

http://dx.doi.org/10.1016/j.cognition.2017.09.006
Received 27 March 2015; Received in revised form 3 September 2017; Accepted 11 September 2017

broader implications for theories of learning in uncertain situations.

Here we restrict our investigation of this question to words that are used ostensively as labels for concrete objects, and how learners determine what the referent object is. We examine one proposed solution to the problem of referential ambiguity: *Cross-situational learning*. Cross-situational learning encompasses the idea that, while there may be an unlimited number of possible referents of a label in any given situation, across situations, the intended referent will most often be present and the other entities that contribute to the referential uncertainty will vary. If learners can track the invariance, then they could link the label to the invariant entity. The potential importance of cross-situational information for at least some word classes—e.g., concrete nouns—is recognized in a variety of theoretical approaches (Blythe, Smith, & Smith, 2016; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; McMurray, Horst, & Samuelson, 2012; Pinker, 1989; Siskind, 1996; Yu & Smith, 2007). But theories differ considerably with respect to the proposed learning mechanisms. Some theories hold that learners expect words to be referential (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Waxman & Gelman, 2009). On this view, learners represent words as fundamentally different stimuli compared to other perceptual entities, treating them as symbols that index concepts (Preissler & Carey, 2004). Learners are then motivated (perhaps by innate social predispositions, Akhtar & Tomasello, 2000) to discover what the referent of a novel word is. On this view, word learning is guided by an over-arching hypothesis about the symbolic nature of words and a motivation to determine reference. In contrast, there are theories that either are agnostic or posit no role at all for the idea of reference as a part of the word learning process (Sloutsky & Robinson, 2008; Yu, 2008; Yu & Smith, 2007). Theories that posit no role for the idea of reference in word learning assume that words are treated and processed the same way as other sounds in the learning environment (Sloutsky, 2009; Sloutsky & Robinson, 2008). Under this account, word learning is assumed to be equivalent to tracking the correspondence between visual stimuli and sounds in a multimodal setting. Other theories also argue for the importance of bottom-up statistical mechanisms but are agnostic regarding the role of reference (e.g., Yu, 2008; Yu & Smith, 2007). These accounts focus on the sources of information available in the input that the learner can use for word learning, making no claims regarding the importance of the concept of reference guiding word learning. For example, Yu and Smith (2007) state, "cross-situational learning may go forward non-strategically and automatically, steadily building a reliable lexicon" (Yu & Smith, 2007). In sum, accounts of word learning differ greatly on their commitment to whether learners rely on the concept of reference to guide successful word learning.

In this paper, we investigate the importance of the concept of reference in word learning by varying the referential status of words. While previous studies provided evidence that children as young as 12 months may understand referential properties of words (e.g., Waxman & Markow, 1995), we are aware of no studies that directly examine the importance of reference in word learning. The approach we take here is to manipulate the presence of the referential intent using the cross-situational word learning paradigm (Yu & Smith, 2007), because the task may be easily modified to introduce task-demand ambiguity. By using cover tasks and interference, we created two conditions where the only difference is the assumption learners bring to the task of learning. Under equivalent referential uncertainty, we examined whether word learners will acquire the correct word-to-referent mappings in a cross-situational learning situation with or without knowing that the words are referential. We found that learners encode the identities of the novel words, but fail to learn the word-to-referent mappings, given identical exposure to cross-situational co-occurrence information, when we interfere with their assumptions about reference. We argue that even when the co-occurrence statistics strongly favor one mapping over others, learners only make the mapping when guided by the broader referential hypothesis.

2. General experimental approach

Our experiments follow the basic design of a cross-situational word learning study, with a learning phase followed by a testing phase. In trials of the learning phase, subjects heard a novel word and saw an array of two pictures, each depicting a common object. Similar to other studies (e.g., Yu & Smith, 2007), subjects heard each novel word in multiple trials in which one of the two objects was present across all trials, but the other objects varied. If learners are indeed biased to seek a meaning for a newly heard word, then we assume this bias is present in artificial word learning experiments as well as in natural language acquisition. Thus, a critical manipulation in our experiments is to try to override this default for half of the subjects, and observe the effects. To achieve this, intermingled with the trials just described were trials in which the two-object array was accompanied by a 440 Hz pure tone (henceforth, *beep*). In addition, subjects in Experiments 1–3 were assigned to a *reference* group or a *categorization* group. Both groups were exposed to identical picture pairs, words, and beep tones during the familiarization phase, but differed in their task with respect to the beep and word trials. Subjects in the *categorization* group were told that they needed to categorize which pictures were accompanied by beeps and which were accompanied by words, but the fact that words referred to objects was intentionally omitted. Subjects in the *reference* condition were explicitly given the instruction that the words referred to the depicted objects.

Our predictions were as follows: If word learning is not supported by the knowledge that words are referential, the instructions to learners in the categorization group should not hinder them from learning the word-object association patterns just as well as subjects in the reference group. On the other hand, if word learning occurs, in part, because learners seek a referent for a novel word (i.e., they have the hypothesis that words refer), then overriding this assumption in the categorization condition should hinder word learning. In contrast, the instructions to the reference group should reinforce learners' prior tendencies.

3. Experiment 1

3.1. Methods

3.1.1. Participants

Twenty-four undergraduate students at the University of Southern California were recruited from the Psychology Department subject pool; 12 were randomly assigned to the reference group and 12 to the categorization group (see *Design and Procedure*).

3.1.2. Stimuli

The pictures were 48 black and white simple line drawings depicting common objects, selected from those in Nishimoto, Ueda, Miyawaki, Une and Takahashi (2012). The novel words were 24 monosyllabic and disyllabic non-words, produced by a female native speaker of Standard American English who was unaware of the purpose of the study (Table 1). Words were recorded in a sound attenuating room, and digitized at a sampling rate of 44.1 kHz.

3.1.3. Design and procedure

3.1.3.1. Training phase. The 48 images were divided into 2 sets of 24 objects, half of which were designated as beep-objects and the rest as word-objects. Each word-object was assigned uniquely to one of the 24

Table 1
Novel words used in all experiments.

antow	*choon*	*dess*	*ghen*	*ghire*	*grimpot*
channer	*dap*	*fex*	*ghip*	*glaik*	*jub*
klide	*lifik*	*mirshow*	*plosit*	*puziv*	*sowch*
lartsue	*lowfa*	*nud*	*pooda*	*refton*	*swech*