Cognition 145 (2015) 53–62

Contents lists available at ScienceDirect

Cognition

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

An integrative account of constraints on cross-situational learning

Daniel Yurovsky *, Michael C. Frank

Department of Psychology, Stanford University, United States

ARTICLE INFO

Article history:
Received 5 August 2014
Revised 16 July 2015
Accepted 24 July 2015
Available online 21 August 2015

Keywords:
Statistical learning
Word learning
Language acquisition
Computational models

ABSTRACT

Word-object co-occurrence statistics are a powerful information source for vocabulary learning, but there is considerable debate about how learners actually use them. While some theories hold that learners accumulate graded, statistical evidence about multiple referents for each word, others suggest that they track only a single candidate referent. In two large-scale experiments, we show that neither account is sufficient: Cross-situational learning involves elements of both. Further, the empirical data are captured by a computational model that formalizes how memory and attention interact with co-occurrence tracking. Together, the data and model unify opposing positions in a complex debate and underscore the value of understanding the interaction between computational and algorithmic levels of explanation.

Natural languages are richly structured. From sounds to phonemes to words to referents in the world, statistical regularities characterize the units and their connections at every level. Adults, children, and even infants have been shown to be sensitive to these statistics, leading to a view of language acquisition as a parallel, possibly implicit, process of statistical extraction (Gómez & Gerken, 2000; Saffran, Aslin, & Newport, 1996). Recent experiments across a number of domains, however, show that human statistical learning may be significantly more limited than previously believed (Johnson & Tyler, 2010; Trueswell, Medina, Hafri, & Gleitman, 2013; Yurovsky, Yu, & Smith, 2012).

We focus here on the use of statistical regularities to learn the meanings of concrete nouns (known as cross-situational word learning; Pinker, 1989; Siskind, 1996; Yu & Smith, 2007). Because words' meanings are reflected in the statistics of their use across contexts, learners could discover the meaning of the word "ball" (for instance) by noticing that while it is heard across many ambiguous contexts, it often accompanies play with small, round toys. A growing body of experiments shows that adults, children, and infants are sensitive to such co-occurrence information, and can use it to map words to their referents (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Vlach & Johnson, 2013; Yu & Smith, 2007).

Information about a word's meaning can thus be extracted from the environmental statistics of its use (Frank, Goodman, & Tenenbaum, 2009; Siskind, 1996). But this analysis is posed at what Marr (1982) called the "computational theory" level: dealing only with the nature of the information available to the learner. At the "algorithmic" level—the level of psychological instantiation in the mind of the learner—this idealized statistical computation could be realized in many ways, and the computation human learners actually perform is a topic of significant debate (see e.g., Yu & Smith, 2012).

Do human learners really track and maintain a representation of word-object co-occurrences? Some evidence suggests that humans are indeed gradual, parallel accumulators of statistical regularities about the entire system of word-object co-occurrences, simultaneously acquiring information about multiple candidate referents for the same word (McMurray, Horst, & Samuelson, 2012; Vouloumanos, 2008; Yurovsky, Fricker, Yu, & Smith, 2014). Other evidence suggests that statistical learning is a focused, discrete process in which learners maintain a single hypothesis about the referent of any given word. This referent is either verified by future consistent co-occurrences or instead rejected, "resetting" the learning process (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al., 2013). While both of these algorithmic-level solutions will, in the limit, produce successful word-referent mapping, they will do so at very different rates. In particular, if learners track a only a single referent for each word, it may be necessary to posit additional biases and constraints on learners in order for human-scale lexicons to be learned in human-scale time from the input available to children (Blythe, Smith, & Smith, 2010; Reisenauer, Smith, & Blythe, 2013).

To distinguish between these two accounts, previous experiments exposed learners to words and objects in which co-occurrence frequencies indicated several high-probability referents for the same word. At the group level, participants in these experiments showed gradual learning of multiple referents for the same word (e.g., Vouloumanos, 2008; Yurovsky, Yu, & Smith,

* Corresponding author.
E-mail address: yurovsky@stanford.edu (D. Yurovsky).

http://dx.doi.org/10.1016/j.cognition.2015.07.013
0010-0277/

2013); but gradual, parallel learning curves can be observed at the group level even if individuals are discrete, single-referent learners (Gallistel, Fairhurst, & Balsam, 2004; Medina et al., 2011). Experiments measuring the same learner at multiple points—a stronger test—have produced mixed results. In some cases, learners showed clear evidence of tracking multiple referents for each word, suggesting a distributional approximation mechanism at the algorithmic level (Dautriche & Chemla, 2014; Smith, Smith, & Blythe, 2011; Yurovsky et al., 2013). In other experiments, however, learners appear to track only a single candidate referent, and to restart from scratch if their best guess is wrong (Medina et al., 2011; Trueswell et al., 2013).

These mixed results expose a fundamental gap in our understanding of the mechanisms humans use to encode and track environmental statistics critical for learning language. Evidence for each account is separately compelling, but neither account can explain the evidence used to support the other. Because previous experiments differ along a number of dimensions—e.g., methodology, stimuli, timing, and precision of measurement—it has been difficult to integrate them to understand why cross-situational learning sometimes appear distributional and sometimes appear discrete (for a review, see Yurovsky et al., 2014).

We propose that differences in task difficulty may explain diverging results across experiments. Two salient dimensions vary across previous studies: ambiguity of individual learning instances, and the interval between successive exposures to the same label (Fig. 1). As attentional and memory demands increase, learners may shift from statistical accumulation to single-referent tracking (Smith et al., 2011; Trueswell et al., 2013).

We present a test of this hypothesis, adapting a paradigm first introduced in Bower and Trabasso (1963) to study the information learners store in concept identification. We parametrically manipulated both the ambiguity of individual learning trials and the interval between them and measured multiple-referent tracking at the individual-participant level. Even at the maximum difficulty tested, learners tracked multiple referents for each word; this result constitutes strong evidence against a qualitative shift from statistical accumulation to single-referent tracking. The data also show that learners encode the referents with differing strengths, however, remembering their hypothesized referent much better. Thus, each previous account appears to be partially correct.

To clarify how these two accounts are related, we implemented both single-referent tracking and statistical accumulation as computational models. We also extended these accounts into an integrative model that subsumes both as special cases along a continuum. Only the integrative model accounted for our full dataset.

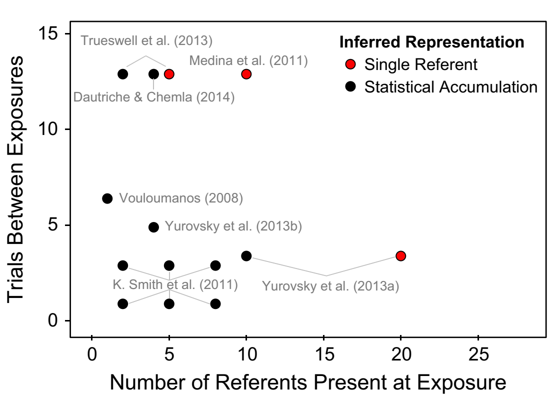

Fig. 1. Results of previous experiments investigating representations for cross-situational learning. These experiments vary along a number of dimensions, but two appear to predict whether multiple-referent tracking is observed: the number of referents present on each trial, and the interval between trials for the referent.

Further, this model was able to make nearly perfect parameter-free predictions for a follow-up experiment that was designed to verify that learners encode mappings rather than individual words and objects. We conclude that cross-situational word learning is best characterized by an integrative account: Learners track both a single target referent and an approximation to the co-occurrence statistics; the strength of this approximation varies with the complexity of the learning environment.

1. Experiment 1

We designed Experiment 1 to estimate learners' memory for both their single best hypothesis about the correct referent of a novel word and their additional statistical knowledge as demands on attention and memory varied. Participants saw a series of individually ambiguous word learning trials in which they heard one novel word, viewed multiple novel objects, and made guesses about which object went with each word. To succeed, participants needed to encode at least one of the objects that co-occurred with a word, remember it until their next encounter with that word, and check whether that same object was again present. If participants encoded exactly one object, they would succeed only when their initial hypothesis was correct. However, the more *additional* objects participants encoded on their first encounter with a word, the greater their likelihood of succeeding even if their initial hypothesis was incorrect.

Rather than allowing chance to determine whether participants held the correct hypothesis on their first exposure to a novel word, the set of novel objects presented on the second exposure to each word was constructed based on participants' choices. On *Same* trials, the participant's hypothesized referent was pitted against a set of novel competitors. In contrast, on *Switch* trials, one of the objects the participant had previously *not* hypothesized was pitted against a set of novel competitors (see Fig. 2). Logically, either a single-referent tracking or a statistical accumulation mechanism will succeed on Same trials. However, only statistical accumulation of information about non-target items can succeed at above-chance levels on Switch trials.

1.1. Method

1.1.1. Participants

Experiment 1 was posted to Amazon Mechanical Turk as a set of Human Intelligence Tasks (HITs) to be completed only by participants with US IP addresses that paid 30 cents each (for a detailed comparison of laboratory and Mechanical Turk studies see Crump, McDonnell, & Gureckis, 2013). Ninety HITs were posted for each of the 16 Referent × Interval conditions for a total of 1440 paid HITs. If a participant completed the experiment more than once, he or she was paid each time, but only data from the first HIT completion was included in the final data set (excluded 180 HITs). In addition, data was excluded from the final sample if participants did not give correct answers for familiar trials (64 HITs, see Design and Procedure). The final sample thus comprised 1196 unique participants, approximately 75 participants per condition (range: 71–81).

1.1.2. Stimuli

Stimuli for the experiment consisted of black and white pictures of familiar and novel objects and audio recordings of familiar and novel words. Pictures of 32 familiar objects spanning a range of categories (e.g. squirrel, truck, tomato, sweater) were drawn from the set constructed by Snodgrass and Vanderwart (1980). Pictures of distinct but difficult to name novel objects were drawn from the set of 140 first used in Kanwisher, Woods, Iacoboni, and