Journal of Memory and Language 99 (2018) 62–73

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The company objects keep: Linking referents together during cross-situational word learning

Martin Zettersten [a,*], Erica Wojcik [b], Viridiana L. Benitez [c], Jenny Saffran [a]

[a] University of Wisconsin-Madison, 1202 W Johnson Street, Madison, WI 53706, United States
[b] Skidmore College, 815 North Broadway Street, Saratoga Springs, NY 12866, United States
[c] Arizona State University, 950 S. McAllister Ave, Tempe, AZ 85287, United States

ARTICLE INFO

Article history:
Received 25 May 2017
Revision received 20 October 2017

Keywords:
Cross-situational word learning
Semantic networks
Language
Memory

ABSTRACT

Learning the meanings of words involves not only linking individual words to referents but also building a network of connections among entities in the world, concepts, and words. Previous studies reveal that infants and adults track the statistical co-occurrence of labels and objects across multiple ambiguous training instances to learn words. However, it is less clear whether, given distributional or attentional cues, learners also encode associations among the novel objects. We investigated the consequences of two types of cues that highlighted object-object links in a cross-situational word learning task: distributional structure – how frequently the referents of novel words occurred together – and visual context – whether the referents were seen on matching backgrounds. Across three experiments, we found that in addition to learning novel words, adults formed connections between frequently co-occurring objects. These findings indicate that learners exploit statistical regularities to form multiple types of associations during word learning.

Introduction

One of the central problems faced by observers attempting to learn the words of a novel language is referential ambiguity (Quine, 1960). When a learner hears a novel word, it is likely that a host of candidate referents will be available in the visual environment. Most investigations focused on this problem ask how learners eliminate competing referents to successfully map a label to a single referent (Medina, Snedeker, Trueswell, & Gleitman, 2011; Smith & Yu, 2008; Trueswell, Medina, Hafri, & Gleitman, 2013; Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014). However, learning a word involves more than forming a mapping between a label and an isolated entity. Learners also encode expectations about the types of objects with which a referent is likely to co-occur, and where the word is likely to be encountered (Boyce, Pollatsek, & Rayner, 1989; Miller, 1999; Roy, Frank, Decamp, Miller, & Roy, 2015; Samuelson, Smith, Perry, & Spencer, 2011; Smith, Suanda, & Yu, 2014). While referential ambiguity presents a hurdle for learning label-object mappings, it also provides an opportunity to learn useful information about the contextual structure of the environment, such as which objects are related to one another.

Learners have many strategies at their disposal for solving the problem of referential ambiguity (Akhtar & Tomasello, 2000; Baldwin, 1993; Markman & Wachtel, 1988; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith & Thelen, 2003). One proposed strategy entails cross-situational word learning (Smith & Yu, 2008; Yu & Smith, 2007). Although any single encounter with the word "tomato" may be referentially ambiguous, only one consistently occurring entity will emerge as the word's most likely referent across multiple encounters with "tomato" (i.e., a round, squishy, and savory fruit). There is substantial evidence from cross-situational word learning tasks that both infants (Smith & Yu, 2008; Vlach & Johnson, 2013; Vouloumanos & Werker, 2009; Yu & Smith, 2011) and adults (Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013; Yurovsky et al., 2014) can successfully map labels to objects across multiple ambiguous training instances by using label-referent co-occurrence statistics.

Notably, most cross-situational word learning studies – with a few exceptions (Dautriche & Chemla, 2014; Chen & Yu, 2017; Kachergis, Yu, & Shiffrin, 2009; Roembke & McMurray, 2016) – lack contextual structure. Associations between label-object pairs are the only reliable patterns; relationships among other elements, e.g. between the objects themselves, are intentionally minimized. In natural learning environments, however, any individual instance of a label is immersed in rich contextual information, such as related nouns and verbs, related objects in the environment, or

* Corresponding author.
E-mail address: zettersten@wisc.edu (M. Zettersten).

https://doi.org/10.1016/j.jml.2017.11.001
0749-596X/

visual scenes that connect word utterances across encounters (Hills, 2013; Hills, Maouene, Riordan, & Smith, 2010; Roy et al., 2015). These kinds of contextual structure highlight connections between entities in the environment, such as the associations among objects. Forming connections between related objects, such as tomatoes and lettuce, is crucial to building semantic knowledge (Landauer & Dumais, 1997; Sadeghi, McClelland, & Hoffman, 2015). In the current study, we focus specifically on whether and how learners encode associations between objects from the structure implicit in cross-situational word learning tasks.

Multiple sources of information may lead learners to form object-object associations. One type of cue is the regularity with which objects co-occur in the world (Bar, 2004). Objects are not randomly distributed in the environment, but instead occur in schema-based clusters. When learners hear the word "tomato", they are more likely to be in the presence of some items (e.g., lettuce, onions, and cucumbers) than others (e.g., soccer balls, cleats, and socks). In this sense, the distribution of object co-occurrences in a learner's environment is *skewed* rather than *uniform*: some objects are more likely to occur in each other's company than others.

Another contextual cue that can help learners link objects together is the presence of a visual context shared across similar locations or scenes (Oliva & Torralba, 2007). A similar visual context can link individual objects that are spatially or temporally distant. For example, while tomatoes and lettuce are objects that are sometimes seen together, they also are often seen within the same prototypical visual context: e.g., they both may often appear on a kitchen counter.[1] Shared visual context may aid in linking objects by creating contextual expectations and by guiding attention towards objects in similar contexts: Regularities between the occurrence of objects (such as tomatoes and lettuce) and visual contexts (e.g., the kitchen counter) influences object recognition, such that specific objects come to be linked to specific visual environments (Brockmole, Castelhano, & Henderson, 2006; Oliva & Torralba, 2007; Vlach & Sandhofer, 2011). The kitchen counter therefore begins to activate expectations for both tomatoes and lettuce, and may act as a cue to link objects across different encounters. Simultaneously, shared visual context may also guide attention to objects occurring within a similar visual context. For example, noticing tomatoes on one end of the kitchen counter and lettuce on the other may lead a learner to recognize a relation between the two. Both the co-occurrence of objects and shared visual context are features of the word learning environment that may influence learners' ability to track meaningful links between objects (e.g., Roy et al., 2015; Vlach & Sandhofer, 2011).

Does tracking object-object links help or hurt word learning? On one hand, the fact that tomatoes and lettuce often occur together and frequently share a similar visual context may make the task of word learning even more difficult: the referent for "tomato" may be harder to disambiguate, particularly from the referent for "lettuce", since the two objects frequently co-occur in the presence of each label. A more uniform distribution of potential referents, and more distinctive or variable visual contexts, by contrast, may help the target referent emerge as the most consistent signal across multiple noisy contexts. On the other hand, learning words involves not just learning label-object mappings, but also forming expectations about the contexts in which words occur (Miller, 1999; Saji et al., 2011). From this perspective, it may be useful for a learner to notice regularities beyond a single label-object mapping. Each labeling event is also an opportunity to learn about the company objects (and their labels) keep.

Previous research suggests that adults use information about the relationships between objects to map objects to novel labels. Specifically, learners can use object-object relations to disambiguate the kinds of objects a label might refer to, such as the fact that a label occurred with animal exemplars rather than items from another category (Dautriche & Chemla, 2014). Other studies have shown that skewed distributions in the frequency with which objects co-occur, as well as thematic groupings among co-occurring objects, can influence how adults learn novel label-object mappings (Chen & Yu, 2017; Roembke & McMurray, 2016; Kachergis et al., 2009). However, word learning moments provide opportunities to not only learn about label-object mappings, but also to learn about the relation between entities occurring in the same context. Exploring the set of candidate referents for a word may lead learners to extract contextual regularities, such as which objects often go together. Furthermore, how learners track these additional regularities may affect how they track the label-object mappings. In the current studies, we assessed adults' ability to form associations between objects in addition to learning the referents of novel words, in the absence of explicit instruction to do so.

In each of the following studies, we asked what adults learn about novel object-object associations as they are engaged in cross-situational word learning and how learning these object-object associations affects word learning. On each trial, learners were presented with one word and four novel objects and were asked to pick the object to which the word referred. The correct word-referent pairings were ambiguous within individual trials, but were disambiguated when word-referent pairs were aggregated across trials, as in the typical cross-situational word-learning task design. No feedback was provided during the training trials. During the test trials, we assessed learning of the relationships between objects.

We exposed learners to object-object links during the learning phase in two ways: by manipulating how frequently specific objects occurred together, and by providing a visual context cue that was identical for pairs of objects. We manipulated the distribution of object co-occurrences by creating two types of object co-occurrence distributions: A *uniform* distribution, that is, a condition in which objects co-occurred equally with each other (as in traditional cross-situational word learning studies), and a *skewed* distribution, where each object occurred more frequently with one particular object than with other objects. We manipulated the presence of a visual context cue, a unique background that was identical for some objects but not others, to highlight the links between objects.

In Experiment 1, we tested whether adults could learn object-object structure when both co-occurrences and shared visual context cues highlighted these links. Adults were presented with a skewed distribution in which pairs of objects occurred frequently together and shared identical background images (skewed distribution and visual cue). We assessed adults' learning of both word-object mappings and object-object connections. Experiment 1 was designed to provide a first measure of whether adults can track both word-object links and object-object links when they are highlighted by distributional and visual context cues. In the subsequent experiments, we assessed the distinct contributions of co-occurrence and visual context cues to encoding contextual structure. In Experiment 2, we asked whether the visual context cue alone (uniform distribution and visual cue) was sufficient for participants to learn object-object connections. In Experiment 3, we asked whether the object co-occurrences alone (skewed distribution and no visual cue) were sufficient for participants to learn object-object connections. In all experiments, we assessed whether encoding contextual structure affected word learning. If tracking

[1] Besides providing a visual cue, kitchen counters also evoke a host of semantic and thematic information for a learner that are relevant to learning associations between objects and label-object associations (see Chen & Yu, 2017). Our goal in the current experiments was to isolate the role of visual cues in the absence of semantic information.