Journal of Memory and Language 84 (2015) 108–127

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Generalizing beyond the input: The functions of the constructions matter

Florent Perek *, Adele E. Goldberg

Princeton University, Department of Psychology, Peretsman-Scully Hall, Princeton, NJ 08540, United States

ARTICLE INFO

Article history:
Received 4 June 2014
revision received 27 April 2015
Available online 16 June 2015

Keywords:
Language acquisition
Artificial language learning
Novel construction learning
Statistical learning
Argument structure
Generalization

ABSTRACT

A growing emphasis on statistics in language learning raises the question of whether and when speakers use language in ways that go beyond the statistical regularities in the input. In this study, two groups were exposed to six novel verbs and two novel word order constructions that differed in function: one construction but not the other was exclusively used with pronoun undergoers. The distributional structure of the input was manipulated between groups according to whether each verb was used exclusively in one or the other construction (the lexicalist condition), or whether a minority of verbs was witnessed in both constructions (the alternating condition). Production and judgments results demonstrate that participants tended to generalize the constructions for use in appropriate discourse contexts, ignoring evidence of verb-specific behavior, especially in the alternating condition. Our results suggest that construction learning involves an interaction of witnessed usage together with the functions of the constructions involved.

Introduction

There is a growing body of research demonstrating that children and adults are acutely sensitive to the statistical properties of the language that they witness, insofar as a great deal of "item-specific" statistical information about particular words is recorded. In fact, the token frequencies of words and phrases play a key role in a number of linguistic processes (e.g., Bybee, 2010; Ellis, 2002; Gibson, Schutze, & Salomon, 1996; Gries & Divjak, 2012). For example, frequent subject auxiliary combinations are more likely to be produced earlier by children than less frequent combinations (Dąbrowska & Lieven, 2005). Sentences tend to be comprehended more quickly when individual verbs appear with complements that are statistically more likely (Ford, Bresnan, & Kaplan, 1982; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994). More frequent combinations of words are more likely to be reduced and/or grammaticalized (Bybee & Hopper, 2001; Gahl & Garnsey, 2004; Kuperman & Bresnan, 2012), are processed faster (Arnon & Snider, 2010; Gathercole & Baddeley, 1993), and are repeated faster and more accurately, both by children and adults (Bannard & Matthews, 2008; Bod, 1998).

Statistical distributional information is recognized as a rich source of evidence available to human and machine learners alike. Transitional probabilities, known to be tracked even by young infants (Saffran, Aslin, & Newport, 1996), can be used to learn higher order generalizations akin to phrase structure rules (Saffran, 2001, 2003). With the wide availability of large corpora, researchers are using co-occurrence statistics as a way of discovering semantic structure as well as formal regularities (Baroni & Lenci, 2010; Landauer, Foltz, & Laham, 1998; Lund & Burgess, 1996; Mintz, Newport, & Bever, 2002). It has been demonstrated that a combination of type and token frequencies in

* Corresponding author at: University of Basel, English Department, Nadelberg 6, CH-4051 Basel, Switzerland.
E-mail address: florent.perek@unibas.ch (F. Perek).

http://dx.doi.org/10.1016/j.jml.2015.04.006
0749-596X/

the input plays an important role in whether learners generalize a novel word or novel grammatical construction beyond their exposure, productively applying it to new instances (e.g., Barðdal, 2008; Casenhiser & Goldberg, 2005; Suttle & Goldberg, 2011; Wonnacott, Boyd, Thompson, & Goldberg, 2012; Wonnacott, Newport, & Tanenhaus, 2008; Xu & Tenenbaum, 2007).

In an important study that inspired the present work, Wonnacott et al. (2008) found that the overall statistics of an artificial language plays a role in how individual items are treated. Artificial language learning experiments involve having participants learn a miniature language by exposing them to a set of novel phrases or sentences that are paired with some sort of interpretation (e.g., Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004; Culbertson, Smolensky, & Legendre, 2012; Wonnacott et al., 2008). Wonnacott et al. demonstrated that adult learners exposed to a "lexicalist" language in which most verbs appeared in only one construction behaved conservatively, avoiding extending verbs for use in a different construction; on the other hand, learners exposed to a "generalist" language in which the majority of verbs alternated, appearing in both constructions, readily assumed that all verbs alternated. Wonnacott (2011) is a similar study that has replicated the basic findings with children.

These sorts of input-driven findings may seem to lead to the conclusion that learners acquire their knowledge of language from simply gleaning statistical regularities from the language input. Is it possible that language learning is *wholly* a process of learning various statistical regularities in the input (e.g., Taylor, 2012)? Weighing against this conclusion is a range of findings that indicate that learners bring to the task of language learning certain biases that help shape what is learned. While certain domain-specific "substantive" biases have been proposed (e.g., Culbertson et al., 2012; but see Goldberg, 2013), other biases have been argued to emerge from the communicative function of language (Hawkins, 1994, 2004, 2014; Jaeger, 2010; Levy & Jaeger, 2007; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Piantadosi, Tily, & Gibson, 2012), from domain general constraints on working memory (Fedorenko, Gibson, & Rohde, 2006; Gathercole & Baddeley, 1993; Hudson Kam & Newport, 2005), for a preference for simplicity (Culbertson & Newport, 2015), or from rational inductive processes (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Perfors, Tenenbaum, & Wonnacott, 2010). It is also well-established that the meanings of words play a role in constraining their distributions and *vice versa*, insofar as semantically related words tend to occur in similar distributional contexts (Arunachalam & Waxman, 2015; Fisher, Gleitman, & Gleitman, 1991; Scott & Fisher, 2009; Waxman, Lidz, Braun, & Lavin, 2009). Somewhat less emphasized have been constraints that emerge from the function of particular constructions (but see e.g., Ambridge & Goldberg, 2008; Ambridge, Pine, Rowland, & Young, 2008; Bybee, 1985; Lakoff, 1987; Langacker, 1987). This is the focus of the present work.

We begin by taking a closer look at the Wonnacott et al. (2008) experiments that had shown that the statistical properties of an artificial language as a whole determined how individual words were used. Participants were taught five novel nouns and 12 novel verbs and were then exposed, over a five-day period, to a language that contained two constructions with different word orders, VSO (Verb Subject Object) and VOS-ka (Verb Object Subject followed by a particle, *ka*). In two experiments, the proportion of verbs that occurred in either construction was varied. Relevantly to our purposes, in their first experiment, a third of the verbs (4) occurred only in the VSO construction, a third of the verbs occurred only in the VOS-ka construction, and the final third appeared in both constructions with equal probability; in this case, participants tended to be lexically conservative, preferring to produce and expecting to hear the one-construction verbs in the construction that they had witnessed those verbs in.[1] The second experiment included twice the number of alternating verbs (two thirds: 8 verbs) as non-alternating verbs (2 verbs were witnessed only in VSO and 2 only in VOS-ka); in this case, there was a much stronger tendency to use all of the verbs in both constructions. Thus learners made use of not only the behavior of individual verbs, but also more general patterns in the input. They appear to implicitly assume roughly, "if few verbs alternate, I will be conservative and only use verbs as I have witnessed them; but if most verbs alternate, perhaps all of the verbs alternate."

A final experiment compared contrasting inputs. One group witnessed a completely alternating language in that all 8 verbs alternated (in a ratio of 7-1 in favor of the VOS-ka construction over the VSO construction). The other group witnessed a completely lexicalist language: 7 verbs appeared only in the VOS-ka construction and 1 verb appeared only in the VSO construction. Subjects learning the alternating language alternated at roughly the same 7-1 rate, even for novel verbs. The learners of the lexicalist language were lexically conservative, using the VOS-ka verbs in that pattern and the VSO verb in its pattern. In other words, in the absence of alternating verbs in the input, learners tended to assume that no verbs could alternate.

It is important to note that the two constructions used in the Wonnacott et al. (2008) experiments were interchangeable, in that there was no discernible difference in their meanings or discourse functions. This situation rarely occurs in natural languages; whenever there exist verbs that alternate between two constructions, there is almost always a functional difference between the constructions. If the constructions do not differ in terms of truth conditions, then they involve a distinction in terms of construal, information structure, pragmatics, register, or dialect (e.g., Bolinger, 1968; Goldberg, 2004; Langacker, 1987). For example, the English double-object construction (e.g., *She gave him a book*) and *to*-dative (e.g., *She gave a book to him*) are a classic case of an alternation, in that many verbs

[1] Wonnacott et al. (2008) also varied the token frequency of different verbs, and found entrenchment effects: the more frequent verbs were less likely to be used in a construction they had not been witnessed in, and were judged to be less acceptable. As the focus of the present work was on the functions of constructions, we did not include a token frequency manipulation.