Journal of Memory and Language 95 (2017) 36–48

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language

Elizabeth Wonnacott [a,*], Helen Brown [b], Kate Nation [c]

[a] Division of Psychology and Language Sciences, University College London, UK
[b] Department of Psychology, University of Warwick, UK
[c] Department of Experimental Psychology, University of Oxford, UK

ARTICLE INFO

Article history:
Received 23 September 2016
revision received 18 January 2017

Keywords:
Language acquisition
Statistical learning
Overgeneralization

ABSTRACT

Successful language acquisition requires both generalization and lexically based learning. Previous research suggests that this is achieved, at least in part, by tracking distributional statistics at and above the level of lexical items. We explored this learning using a semi-artificial language learning paradigm with 6-year-olds and adults, looking at learning of co-occurrence relationships between (meaningless) particles and English nouns. Both age groups showed stronger lexical learning (and less generalization) given "skewed" languages where a majority particle co-occurred with most nouns. In addition, adults, but not children, were affected by overall lexicality, showing weaker lexical learning (more generalization) when some input nouns were seen to alternate (i.e. occur with both particles). The results suggest that restricting generalization is affected by distributional statistics above the level of words/bigrams. Findings are discussed within the framework offered by models capturing generalization as rational inference, namely hierarchical-Bayesian and simplicity-based models.

Introduction

A classic problem for theories of language acquisition is how learners avoid overgeneralization in the face of an ability to generalize. An example is our knowledge of restrictions on novel combinations of verbs and argument structures, as in the use of "*carry*" in the double-object dative e.g., *"*Carry me that*". Children go through a stage of producing overgeneralizations, yet eventually learn that certain combinations of verbs and structures are restricted. This "paradox" (Baker, 1979) has received a good deal of attention in the literature. Broadly, two different classes of solution have been proposed, one emphasizing increasing knowledge of the semantics of words and constructions (e.g., Pinker, 1989) which eventually provide constraints that block overgeneralizations, and one emphasizing the use of distributional statistics to make inferences about which generalizations are permissible (e.g., Braine, 1971). There is a growing body of evidence suggesting that generalization is constrained by both types of information and that grammatical learning can be characterized as graded rather than absolute (e.g., Ambridge, Pine, Rowland, Freudenthal, & Chang, 2014). This is consistent with the notion that children acquire probabilistic constraints from input distributions (e.g., Hsu & Chater, 2010; Matthews & Bannard, 2010; Perfors & Wonnacott, 2011; Wonnacott, 2011; Wonnacott, Boyd, Thomson, & Goldberg, 2012).

To inform theory, it is important for experimental work to identify what types of distributional information influence learning and generalization and under which conditions. Here we assess children and adults' sensitivity to a particular distributional property which we term skew. Specifically, we ask whether it is easier to learn arbitrary, lexically based restrictions when structures are not evenly distributed across lexical items (i.e., more words occur with one structure than the other). We also probe the finding from earlier work (Wonnacott, 2011; Wonnacott, Newport, & Tanenhaus, 2008) that it is easier to learn lexical restrictions given broader experience of lexical restrictions within the language.

Artificial language learning provides an ideal tool for exploring learners' sensitivity to different input statistics in isolation of other cues (e.g., semantic, phonological). Wonnacott et al. (2008) took this approach in a series of experiments with adult learners. The input languages incorporated two competing transitive structures and were constructed so that some verbs alternated between structures, but others occurred in just one structure (an arbitrary restriction, since the constructions were synonymous and there were no semantic or phonological cues to verb distribution). Different input sets were used in different learning conditions such that the distributional relationship between verbs and structures was

* Corresponding author at: Division of Psychology and Language Sciences UCL, University College London, 2 Wakefield Street, London WC1N 1PF, UK.
E-mail address: e.wonnacott@ucl.ac.uk (E. Wonnacott).

http://dx.doi.org/10.1016/j.jml.2017.01.005
0749-596X/

manipulated. Participants were given production and judgment tests after exposure to one of these input sets, and generalization was deemed to have occurred when they produced, or accepted as grammatical, an unattested verb-structure combination. Generalization was found to be affected by the distributional statistics of the learner's input. One factor was verb frequency: verbs frequently encountered in one structure were less likely to be generalized to the other. Importantly, however, participants' learning of verb-structure pairings was affected not only by the frequency of those pairings but also by their more general experience of the language being learned. The likelihood of generalization was influenced by the learners' broader experience of alternation across the input: verbs which had only occurred in one construction were more likely to be generalized to the alternate construction if the learner had experienced more alternating verbs in the input.

Wonnacott (2011) used an adapted learning paradigm to replicate aspects of these findings with 6-year-old children. There are relatively few artificial language learning experiments with children beyond infancy (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Hudson Kam & Newport, 2005, 2009; Wonnacott, 2011; Wonnacott et al., 2012). Those that have been conducted indicate that children's learning is substantially slower than that of adults. For example, Wonnacott et al. (2012) found that after three days of training on a single novel verb-argument construction, children produced the structure with correct linking of word order to thematic roles on only 57% of trials, while adults were at ceiling.

The observation that children's learning is slower than adults has implications for experimental design. Unfortunately however, it is not straightforward to simply add additional exposures to compensate for the slower rate of learning. Children can only tolerate short experimental sessions, and schools cannot generally accommodate additional sessions to mitigate this. It is thus necessary to design artificial languages where the "baseline" structures can be acquired relatively quickly. Given these constraints, in order to be able to directly focus on the balance between generalization and lexically specific constraints given relatively little exposure, Wonnacott (2011) used a learning paradigm where the critical relationships were between nouns and meaningless words referred to as "particles", rather than verbs and verb constructions. To facilitate learning, the languages used novel particles but familiar English nouns. This simpler paradigm allowed the same types of statistical manipulations as in Wonnacott et al. (2008) to be explored, with languages containing both alternating nouns (i.e., nouns which occurred with both particles) and nouns restricted to occur with just one particle. A production test was used to probe generalization.

In line with the previous effects of verb frequency in adults, noun frequency played a role, with more generalization to the non-occurring particle for low frequency nouns. Again, however, generalization was also affected by learners' more general experience of the language being learned. Most relevant to the current work, Experiment 1 compared the learning of *minimal-exposure items* in different language contexts. Each of the two *minimal-exposure items* occurred only with one of the two particles, and both were low frequency (presented four times each). The question was whether learners would restrict their usage of *minimal-exposure items* to the particle with which it had occurred in the four exposures, or generalize and extend it to the other particle. From the perspective of individual lexical frequency, four exposures is a very small sample and learners might therefore be expected to ignore this item-specific input and generalize. Importantly, however, these items were introduced later in the experiment, *after* the participants had been exposed to language input containing other nouns. How these *minimal-exposure items* were treated depended critically upon the input to which the children had been previously exposed. Those previously exposed to an input language where each noun occurred with just one of the two particles (dubbed the *lexicalist* language) were more likely to avoid generalizing with the *minimal-exposure nouns*, treating them as restricted to occur with the one particle with which they were attested. In contrast, learners who had been exposed to a language where all verbs alternated (dubbed the *generalist* language) treated *minimal-exposure nouns* as alternating. Thus children's learning of the restrictions on particular nouns appeared to be affected by their more general learning of how nouns tended to behave across the whole language.

An additional factor explored in the same experiment, and using the same input languages, was whether children could pick up on the statistical prevalence of the particles in the language overall. To this end, in both languages there was a 3:1 bias for one particle, achieved in the *lexicalist* language by having three nouns occur with particle1, and one noun with particle2, and in the *generalist* language by having the 4 alternating nouns each have a bias to occur three times more often with particle1. Testing with entirely novel nouns revealed that children exposed to both *lexicalist* and *generalist* languages had learned the particle1 bias – i.e., they were more likely to generalize that particle. In addition, children in the *generalist* condition were more likely to overgeneralize particle1 with the *minimal-exposure nouns*.

Perfors, Tenenbaum and Wonnacott (2010; Wonnacott & Perfors, 2009) demonstrated that this pattern of learning is in line with the predictions of a hierarchical Bayesian model. This domain general model was originally developed by Kemp, Perfors, and Tenenbaum (2007), who applied it to a set of cognitive learning problems (e.g., acquisition of the "shape bias" in word learning). It is characterized by an ability to track statistical distributions at multiple levels of abstraction (in our work, the distribution of particles used with particular nouns and the language-wide distribution of particles), and to make inferences about the extent to which these levels provide a good indicator of future behavior. This is achieved via the formation of "overhypotheses" about a particular dataset. For example, when trained on the *lexicalist* language, the model formed an "overhypothesis" to the effect that the usage of particles was highly consistent for particular nouns, whereas when trained on the *generalist* language it formed the "overhypothesis" that noun identity and particle usage were unrelated. These "overhypotheses" led to the model showing the same difference in the learning of *minimal-exposure items* as human learners, i.e., greater learning of the associations between these items and their attested particles in the *lexicalist* than *generalist* language. The model also mimicked human performance in showing greater generalization with the more frequent of the two particles/structures, both with novel items and with the *minimal-exposure items* in the *generalist* language. This is due to the fact that it tracked their distribution across the whole language.

The current work builds on previous work by focusing on a property of the lexicalist-input sets used by Wonnacott (2011): the *skewed* distribution of particles across *input nouns*. This skew was originally included to explore the learning of language-wide patterns of particle usage. Potentially however, skew might in itself be an aid to lexical learning. Skewed distributions are common in natural languages. For example, constructions tend to occur more frequently with a single verb (e.g. the double-object [DO]-dative construction occurs more with "give" – "*he gave her the present*" - than with any other construction, and this distribution may benefit learning of its meaning; Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004). Another type of skewed distribution is common in grammatical systems where there are alternative forms serving the same function. In this situation, it is often the case that there is one particular form which is used with the majority of lexical items (e.g. the regular English plural *–s*) while other forms are used with a minority of lexical items (e.g. English plural exception forms such as *feet* and *children*).