Journal of Memory and Language 65 (2011) 1–14

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Balancing generalization and lexical conservatism: An artificial language study with child learners

Elizabeth Wonnacott *

The University of Oxford, UK

ARTICLE INFO

Article history:
Received 17 September 2010
revision received 22 February 2011
Available online 1 April 2011

Keywords:
Language acquisition
Statistical learning
Artificial language learning
Over-generalization
Baker's Paradox

ABSTRACT

Successful language acquisition involves generalization, but learners must balance this against the acquisition of lexical constraints. Such learning has been considered problematic for theories of acquisition: if learners generalize abstract patterns to new words, how do they learn lexically-based exceptions? One approach claims that learners use distributional statistics to make inferences about when generalization is appropriate, a hypothesis which has recently received support from Artificial Language Learning experiments with adult learners (Wonnacott, Newport, & Tanenhaus, 2008). Since adult and child language learning may be different (Hudson Kam & Newport, 2005), it is essential to extend these results to child learners. In the current work, four groups of children (6 years) were each exposed to one of four semi-artificial languages. The results demonstrate that children are sensitive to linguistic distributions at and above the level of particular lexical items, and that these statistics influence the balance between generalization and lexical conservatism. The data are in line with an approach which models generalization as rational inference and in particular with the predictions of the domain general hierarchical Bayesian model developed in Kemp, Perfors & Tenenbaum, 2006. This suggests that such models have relevance for theories of language acquisition.

Introduction

Successful language acquisition requires an ability to generalize, but learners must balance this against the acquisition of exceptions. For example, English native speakers regularly combine adjectives and nouns in novel ways, yet there are some arbitrary restrictions on usage: *strong winds* may be *high winds*, but *strong breezes* are not? *high breezes.* These preferences do not arise from any obvious semantic constraints – rather they appear to rely on knowledge that is arbitrary and lexically-based. Another example is verb sub-categorization preferences. For example, a number of English verbs can occur in both of two near synonymous dative structures: the prepositional-dative, e.g. *I told the story to Ben*, and the double-object-dative, e.g. *I told Ben the story*, yet others are restricted to occur only the prepositional-dative, as in *I explained the story to him, *I explained him the story*. Again, these lexical exceptions are not obviously explained by semantic constraints.

Children (at least from age 3, see Tomasello, 2000) are able to generalize and use words in new ways, as seen both in overgeneralization errors (*Don't say me that!*) and the usage of nonce words in new constructions in experiments (e.g. Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989). How then do they learn that certain usages are impermissible? This quandary, known as *Baker's Paradox* (Baker, 1979), has been considered central to theories of language acquisition. One solution attempts to show that apparent exceptions actually arise from more general regularities based on the semantic or phonological properties of words (Pinker, 1989). However, although such factors influence generalization in adults and older children (e.g. Gropen et al., 1989), young children may not be sensitive to the requisite conditioning criteria (Brooks & Tomasello, 1999;

* Address: Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK. Fax: +44 (0) 1865 310447.
E-mail address: elizabeth.wonnacott@psy.ox.ac.uk

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.03.001

Ambridge, Pine, Rowland, & Young, 2008). Their role in development is thus unclear. Moreover, various researchers have disputed the claim that these cues can fully determine verb-syntax (e.g. Braine & Brooks, 1995), suggesting that learners must be capable of learning arbitrary, lexically-specified exceptions.

An alternative approach, originating with Braine (1971), suggests that learners are sensitive to lexically based distributional statistics and use this information to make inferences about when generalization is and is not appropriate. This concurs with approaches to language acquisition that emphasize the role of statistical learning processes (Elman, 1990; Rumelhart & McClelland, 1986; Saffran, Aslin & Newport, 1996). There is evidence that children and adults are more likely to allow novel generalizations with low frequency lexical items. For example, children are more likely to over-generalize with low frequency than high-frequency verbs, judging "*He came me to school*" to be worse than "*He arrived me to school*" (Ambridge et al., 2007; Brooks, Tomasello, Dodson, & Lewis, 1999; Theakston, 2004). This has been explained in terms of *Entrenchment* (Braine & Brooks, 1995) or *Statistical Pre-emption* (Goldberg, 2005)[1]: frequently encountering verbs with alternative constructions leads to reluctance to generalize to a new construction.

In addition to the frequency of lexical items, higher-level statistics may also affect generalization. Goldberg (2005) argued that high-frequency verb argument structures such as the transitive are more likely to be generalized to new verbs. Corroborating evidence comes from the sentence processing literature. When reading or listening to language, we make predictions about upcoming sentence structure which concur with the verb's distributional history (e.g. Trueswell, Tanenhaus, & Kello, 1993), but may also be influenced by "higher-level" biases – for example to interpret post-verbal nouns as direct objects even for intransitive verbs (Mitchell, 1987). This latter effect may be due to the greater frequency of the transitive structure across the language (Juliano & Tanenhaus, 1994). Such findings about language *processing* suggest that language *learning* involves accumulating statistics at both a lexically specific and more generalized level.

A recent study by Wonnacott et al. (2008) provided direct evidence that adult learners can use both lexically specific and higher-level statistics in constraining generalization. In order to establish that effects were driven by distributional patterns at and above the lexical level, rather than semantic or phonological motivations, an Artificial Language Learning methodology was used, i.e. participants were exposed to experimenter-created miniature languages and tested to see when they generalized. Specifically, the miniature languages incorporated two competing transitive structures; generalization occurred when a participant used a verb with a structure with which it did not occur in the input. In addition to individual verb frequency, two 'higher-level' statistical factors were found to influence the usage of a verb in a novel construction (a) the frequency of the structure across the language: more generalization with a higher frequency structure (b) the distribution of verb types across the language: if *most* verbs across their input language had occurred in both of the two structures (so called *alternating* verbs), learners were more likely to generalize verbs from one construction to the other. These effects were particularly clear with what Wonnacott et al. called *minimal-exposure* verbs. These were verbs which occurred in only one of the two structures and with very few exposures (four). Importantly, they were presented to learners only *after* they had been previously exposed to a large amount of language input involving other verbs. Wonnacott et al. asked whether learners would restrict their usage of these verbs to the structure in which it had been encountered (lexical conservatism), or extend it to the other structure (generalization). From the perspective of individual lexical frequency, four exposures is a very small sample, learners might therefore be expected to ignore this verb-specific input and generalize. In fact, learners' treatment of these verbs depended upon the input to which they had been previously exposed: participants previously exposed to a language where all verbs occurred in just one structure (dubbed the Lexicalist language), showed strong conservatism and little generalization; in contrast, learners exposed to a language where all verbs occurred in both structures (dubbed the Generalist language), generalized those verbs to both structures, particularly generalizing the structure that was of higher frequency across the language.

Wonnacott et al. argued that their learners were taking a *rational* approach to determining when to generalize from minimal evidence, drawing on a theoretical framework provided by Bayesian approaches to cognition. This was formalized by Perfors, Tenenbaum, and Wonnacott (2010) who demonstrated that the data are in line with the predictions of a hierarchical Bayesian model (henceforth HBM). This domain general model had been developed by Kemp, Perfors, and Tenenbaum (2007), who applied it to a distinct set of cognitive learning problems (for example, the problem of acquiring the "shape bias" in word learning), yet it could predict the behavior of the adult artificial language learners. Critically, the model is characterized by an ability to track statistical distributions at multiple levels of abstraction, 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 "over-hypotheses" about a particular data set. For example, when it was trained on the Lexicalist language from Wonnacott et al. (2008), the model formed an "over-hypothesis" to the effect that the usage of constructions was highly consistent for particular verbs, whereas in the Generalist language it formed the over-hypothesis that verb identify and construction usage were unrelated. These over-hypotheses led to the model showing the same difference in the learning of minimal-exposure verbs as human learners, i.e. more learning of the lexical constraints in the Lexicalist than Generalist languages. The model also mimicked human learners in showing greater generalization with the more frequent of the two constructions, due to the fact that it tracked their distribution across the whole language.

[1] The entrenchment and statistical pre-emption hypotheses are subtly different since the latter assumes that novel verb-structure pairings, or other generalizations, are only blocked by encountering near synonymous alternatives. This difference is beyond the scope of the current work since the structures which are generalized in our experiments carry no semantics.