Journal of Memory and Language 92 (2017) 343–358

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Competition between phonological and semantic cues in noun class learning

Jennifer Culbertson *, Annie Gagliardi, Kenny Smith

University of Edinburgh, Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Received 27 April 2016
revision received 15 August 2016

Keywords:
Noun classification
Gender
Artificial language learning
Cue competition
Category learning
Language acquisition

ABSTRACT

Learning noun classification systems, like gender, involves inferring a language-particular set of (often probabilistic) cues to class membership. Previous work has shown that learners rely disproportionately on phonological cues (e.g., Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981). Surprisingly, this occurs even when competing semantic cues are more reliable predictors of class. We investigate two possible explanations for this: first, that phonological cues are more salient to learners than semantic cues, and second that phonological cues are generally available earlier than semantic cues. We show that adult learners' treatment of conflicting cues to noun class in a miniature artificial language depends on *both* cue saliency and early availability. Importantly, learners prioritize earlier-available cues even when they are less salient than competitor cues. Our findings suggest a possible mechanism for children's over-reliance on phonology: children start building their classifications systems very early, when phonological information is available, but word meanings are not.

Introduction

Noun classification systems are grammatical devices, found in many languages, which serve to categorize nouns into classes according to some set of features. They include gender systems made up of a small set of classes (e.g., 'mas culine'/'feminine' as in the Romance languages), larger noun class systems (with up to 10 classes, as in many Bantu languages), and numeral classifier systems (sometimes with hundreds of distinct categories, as in many East Asian languages). These systems differ in their formal realization, the locus of the cues to class (e.g., based on distributional or morphological information, the semantics of noun referents, or phonological properties of nouns themselves) and in the particular cue features which are relevant.

Interestingly, all noun classification systems exploit semantic cues to some extent; there are no known noun classification systems which are based on phonological cues alone (Aikhenvald, 2000). In fact, the set of semantic features used is often similar across languages: natural gender, animacy, and shape are very common (Aikhenvald, 2000; Comrie, 1989; Denny, 1976; Dixon, 1986; Lakoff, 1987; Senft, 2000). That said, many languages exhibit mixed systems, with semantic cues like natural gender along with a set of noun-internal phonological cues which are predictive of class. The particular set of phonological features used varies widely across languages. For example, a prototypical mixed gender system can be seen in French. Natural gender is a highly reliable cue to class: nouns with female gender referents are typically found in one class, while nouns with male gender referents are in another class. Other semantic cues are less reliable, but nevertheless probabilistically cue class; nouns referring to paths or roads are typically feminine, and nouns

* Corresponding author at: Dugald Stewart Building, 3 Charles Street, Edinburgh EH89AD, United Kingdom.
E-mail address: jennifer.culbertson@ed.ac.uk (J. Culbertson).

http://dx.doi.org/10.1016/j.jml.2016.08.001
0749-596X/

whose referents are superordinate categories are typically masculine (Nelson, 2005). Alongside semantics, many phonological features of nouns probabilistically cue class; according to Surridge (1986), there are 34 distinct suffixes which cue one gender or the other in French (see also Lyster, 2006).

This combination of semantic and phonological cues, some more reliable than others, leads to a complex problem for both first and second language learners (Arnon & Ramscar, 2012; Braine, 1987; Carroll, 1999; Frigo & McDonald, 1998; Kempe & Brooks, 2001; Levy, 1988; MacWhinney, 1978; Taraban, 2004; a.o.). Here, we focus on a recurring finding in the literature on first language acquisition of noun classes: the weak role of semantic cues relative to phonological cues. In her classic work on the acquisition of French gender, Karmiloff-Smith (1981) found, across a series of experiments, that children even up to age 10 determine the gender of novel nouns according to their phonological properties (e.g., word endings), rather than exploiting highly reliable semantic cues like natural gender. For example, in one experiment, children (age 3–12) were presented with two pictures of unfamiliar alien characters, one clearly female, the other male. These were given noun labels whose phonology either conflicted with or matched the natural gender of the referents. For example, they might see a male alien and hear the label *podelle,* where *–elle* is a cue to feminine grammatical gender, or a female alien labeled *bicron,* where *–on* is a cue to masculine grammatical gender. The children were subsequently asked to produce these nouns in contexts which required a determiner whose form is determined by grammatical gender (e.g., *le/la*, 'the$_{masc}$/the$_{fem}$'). For example, the experimenter might put an object on one of the pictures, and ask *Qu'est-ce que j'ai fait?* 'What did I do?', and the child might answer *Vous avez mis un crayon sur la podelle* 'You put a pencil on the$_{fem}$ podelle'. Children up to the age of 10 consistently used the determiner which matched the phonological features of the noun, apparently ignoring the semantic cue (e.g., producing *la podelle* for a male alien or *le bicron* for a female alien).

These findings have been reproduced in several other languages including German (Mills, 1985), Spanish (Mariscal, 2009; Perez-Pereira, 1991), Sesotho (Demuth, 2000; Demuth & Ellis, 2008), Russian (Rodina & Westergaard, 2012) and Tsez (Gagliardi, 2012; Gagliardi & Lidz, 2014). Both natural gender and the phonological endings used in Karmiloff-Smith (1981) are highly reliable cues to class in French–even though both are only relevant for a subset of nouns in the language (e.g. natural gender is irrelevant for inanimates, *–elle* and *–on* endings only appear on some nouns), when they are present an ideal observer could use either cue to guess the grammatical gender with high accuracy. Importantly, in the case of Tsez, Gagliardi and Lidz (2014) found that even when phonological cues were statistically *less* reliable than competing semantic cues, young children still preferentially use them to determine class membership. They report that this effect disappears in older children and adults. In contrast to this body of evidence suggesting a preference for phonological over semantic cues, we are aware of only one study which finds that children use semantic cues preferentially (in Icelandic; Mulford, 1985). Interestingly, Mulford suggests this result may be due to the extremely low predictive power of phonological cues in Icelandic (e.g., relative to many such cues in French).

Given what we know about noun classification systems in general, this apparent over-reliance on phonological cues during learning is surprising: semantic cues to class are ubiquitous in language, are in some cases highly reliable, and often pick out very similar properties across languages. By contrast, only some languages make use of phonological cues, and these cues are often less predictive, and highly variable. Why then, would children rely on phonology over semantics? Several distinct mechanisms have recently been proposed to explain this puzzling finding. For example, learners may be actively biased against using external cues like semantics in forming grammatical categories, particularly when internal properties of nouns are available to cue class (Culbertson & Wilson, 2013; Gagliardi, 2012; Gagliardi, Feldman, & Lidz, in press; Gagliardi & Lidz, 2014). Perhaps relatedly, semantic cues may simply be less salient to learners than phonological cues, at least in the context of acquiring noun classification systems (Gagliardi, 2012; Gagliardi et al., in press). Alternatively, learners may have access to phonological properties of nouns before semantic features of their referents (Carroll, 1999; Culbertson & Wilson, 2013; Demuth, 2000; Gagliardi et al., in press; Polinsky & Jackson, 1999), simply because children encounter linguistic forms prior to mapping those forms to their referents. Under this latter explanation, there is no inherent bias against the use of semantic cues (they may even be *more* salient). However, if learners build representations of categories first based on phonology, before they have acquired the relevant word-to-meaning mappings, these initial representations may persist–either because representations initially built on purely distributional properties take some time to incorporate external cues, or because of an asymmetry in the relative amount of data learners have for each type of cue, with more data being available for early-available cues.

Which of these hypotheses is correct has clear implications for theories and models of the acquisition of noun classification systems. Here we conduct the first experimental tests of the proposed mechanisms outlined above, focusing on salience and early availability of cues. We use an artificial language learning paradigm in order to maintain complete control over the reliability and frequency of different cues. Following previous studies of artificial noun class learning (Braine et al., 1990; Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Culbertson & Wilson, 2013; Frigo & McDonald, 1998; Williams, 2004), we use adult learners: the hypotheses we are testing do not specifically distinguish child from adult learners, and if certain cues are more salient than others, we expect they will be for learners of any age. Our laboratory learning task allows us to make some cues available to learners earlier than others, re-creating a learning environment for adults that would be similar to first language acquisition in this key aspect.