Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation

Anna Samara [a],*, Kenny Smith [b], Helen Brown [c], Elizabeth Wonnacott [a]

[a] Division of Psychology and Language Sciences, University College London, United Kingdom
[b] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, United Kingdom
[c] Department of Psychology, University of Warwick, United Kingdom

A B S T R A C T

Languages exhibit sociolinguistic variation, such that adult native speakers condition the usage of linguistic variants on social context, gender, and ethnicity, among other cues. While the existence of this kind of socially conditioned variation is well-established, less is known about how it is acquired. Studies of naturalistic language use by children provide various examples where children's production of sociolinguistic variants appears to be conditioned on similar factors to adults' production, but it is difficult to determine whether this reflects knowledge of sociolinguistic conditioning or systematic differences in the input to children from different social groups. Furthermore, artificial language learning experiments have shown that children have a tendency to eliminate variation, a process which could potentially work against their acquisition of sociolinguistic variation. The current study used a semi-artificial language learning paradigm to investigate learning of the sociolinguistic cue of speaker identity in 6-year-olds and adults. Participants were trained and tested on an artificial language where nouns were obligatorily followed by one of two meaningless particles and were produced by one of two speakers (one male, one female). Particle usage was conditioned deterministically on speaker identity (Experiment 1), probabilistically (Experiment 2), or not at all (Experiment 3). Participants were given tests of production and comprehension. In Experiments 1 and 2, both children and adults successfully acquired the speaker identity cue, although the effect was stronger for adults and in Experiment 1. In addition, in all three experiments, there was evidence of regularization in participants' productions, although the type of regularization differed with age: children showed regularization by boosting the frequency of one particle at the expense of the other, while adults regularized by conditioning particle usage on lexical items. Overall, results demonstrate that children and adults are sensitive to speaker identity cues, an ability which is fundamental to tracking sociolinguistic variation, and that children's well-established tendency to regularize does not prevent them from learning sociolinguistically conditioned variation.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Variation is ubiquitous in natural language and occurs at all levels of analysis, be it phonetic, morphological, syntactic, semantic, or lexical. However, the usage of linguistic variants tends to be *conditioned*, so that variation is rarely, possibly

never, fully unpredictable. A clear example of deterministic conditioning is the regular past tense marker in English (written as –ed), which can be realized as [t], [d], or syllabic [ɪd] (as in *liked*, *loved*, *hated*), with the choice of variant dependent on the phonological features of the final segment of the stem. However, conditioning contexts may also be probabilistic. For example, in English, a final (–t, d) segment in a final cluster is variably deleted (*nest* vs. *nes'*) with the probability of deletion affected by a variety of linguistic factors, including the phonological features of the following segment (e.g., /t/ or /d/ followed by an obstruent is more likely to be deleted than /t/ or /d/ followed by a liquid), morphological class (e.g., final /t/ or /d/ in monomorphemes delete more frequently than weak past tense forms) and the presence of a following pause, as well as social factors such as the speaker's gender and social context. We will refer to patterns of variation involving these latter kinds of social conditioning factors as *sociolinguistic*.

Although the existence of sociolinguistic variation is well-established in adult language, less is known about how children respond to socially conditioned variation during acquisition. For example, can young children pick up on the fact that different speakers use different variants, even when that relationship is probabilistic (as is often the case in adult languages)? How does the learning of socially conditioned variation interact with their well-demonstrated tendency to regularize experimenter-created miniature languages that exhibit fully unpredictable variation? The current paper explores these questions using a statistical learning framework, comparing children's and adults' learning of a form of sociolinguistic variation (variable forms cued by speaker identity) in an artificial language.

1.1. Sociolinguistic variation in adult learners

In his seminal work, Labov (1963) demonstrated that linguistic variation may be subject to probabilistic, extra-linguistic social constraints. He found that the pronunciation of some vowels on Martha's Vineyard (an island off the coast of Massachusetts) was changing from the standard American pronunciation, and that this was affected by social identity: age, occupation, and social group were all important factors, with those who identified most closely as natives of the island using the incoming forms most frequently. Social influences on linguistic choice were further established in Labov's (1966) "New York department store study", which examined differences in the pronunciation of postvocalic /r/ (e.g., as occurring in *fourth*, *floor*) in New York speech. Rhotic varieties (where /r/ is pronounced post-vocalically) were associated with prestige in New York, and were favoured by employees of an upper-middle class shopping center, whereas /r/-dropping was more frequently used by employees in a shopping center with a more working-class clientele. This work also revealed *within-speaker* variation and correlations between language style and context, with rhoticity notably higher in the context of careful or emphatic speech, particularly so among employees catering for a middle-class clientele.

Over the last 50 years, a wide range of sociolinguistic studies have since confirmed Labov's early findings, showing that in adult language, production of phonological and grammatical variation is in accordance with a number of extra-linguistic social factors. For example, there is a large literature demonstrating differences in male and female language use (e.g., Labov, 1966, 2001; Neu, 1980; Trudgill, 1974; Wolfram, 1969). This shows that speakers may associate certain variants with gender and avoid variants they perceive as gender-inappropriate (Cameron, 2005; Eckert & McConnell-Ginet, 1999). Women may also use forms that represent stable social variables (i.e., "standard" forms) in their speech more frequently than men (Labov, 1990). Age (e.g., speech before and after adolescence) and ethnicity (i.e., ethnically marked ways of speaking) have been further shown to exert an influence on linguistic choices (e.g., Eckert, 1997; Foulkes, Docherty, & Watt, 2005; Hoffman & Walker, 2010; Knack, 1991; Mendoza-Denton, 2008).

1.2. Sociolinguistic variation in child language

Traditionally, sociolinguistic research has focused on adult language users rather than addressing questions of acquisition. An early reason was an assumption (originating with Labov, 1964) that child speech is monostylistic early in development, with little sociolinguistic competence before the age of 12. The methodological complications in working with children (e.g., the inadequacy of the sociolinguistic interview for the elicitation of speech in young children, practical difficulties in obtaining corpus of data large enough to locate sociolinguistic variation; Foulkes, Docherty, & Watt, 2001; Roberts, 1994) also proved a barrier. Labov (1989) was among the first to empirically address the question of *when* children first exhibit adult-like command of sociolinguistic variables. He analysed the speech of three Philadelphian children (aged 6, 7, and 9 years) during family interaction, focusing on two sociolinguistic target variables, namely (–t, d) deletion and alternations between /ŋ/ and /ɪn/ in progressive inflection on verbs. In both of the 6- and 7-year-old's speech, there was some evidence that some patterns of variation seen in their parents' speech and their local dialect were present (for example, /t/ and /d/ were hardly ever deleted before pauses), although other constraints were not mastered (for example, unlike adults, they did not show greater likelihood of deletion in certain phonological contexts). While other work has not replicated these precise findings with this variable, the emerging literature confirms that conditioned variation *does* appear in children's speech from early ages, and that the older the children get, the more factors condition variation in their speech (Shin, 2016). Importantly, the exact age of mastery differs across variables and constraints, possibly due to their differential distributional tendencies in the input and their complexity (Shin, 2016). For example, Roberts (1997) studied (-t,-d) deletion in sixteen 3–4 year olds in Philadelphia using an elicitation paradigm. She found that even these young children had acquired some relevant grammatical and phonological constraints (in contrast to Labov's findings); however, patterns of (t- d) deletion did not differ as a function of formal versus informal speech (elicited using different activities), or style of interaction (studied by