From: Kuhl., P. K. & Meltzoff, A. N. (1997)
M. Gopnik (Ed.), *The inheritance and innateness of grammars*
(pp. 7-44). NY: Oxford University Press, 1997.

2

Evolution, Nativism and Learning in the Development of Language and Speech

Patricia K. Kuhl and Andrew N. Meltzoff

Introduction

Infants acquire language like clockwork. Whether a baby is born in Stockholm, Tokyo, Zimbabwe or Seattle, at 3 months of age, a typically developing infant will coo. At about 7 months the baby will babble. By their first birthday, infants will have produced their first words, and by 18 months, 2-word combinations. Children of all cultures know enough about language to carry on an intricate conversation by 3 years of age.

When our own daughter began to produce the "babababa" characteristic of canonical babbling, we were struck by the regularity of its form and the precision of its timing. Having occurred on schedule rather than being accelerated by our ever-constant modelling, we were reminded that the milestones of human language occur at the appointed time regardless of the language in which the child is being reared, the educational background of the infant's parents – and, apparently, regardless of parental prompting or the theories they hold.

Such observations seem to support Chomsky's nativist view that language milestones occur at pre-specified times, as do the eruption of teeth or the onset of puberty. Recent discoveries, however, require revisions to this idea. The emerging view remains strongly nativist, to be sure, but suggests a critical role for language input. The new view provides some insight into how one *particular* language rather than another is acquired. Not only the fact that infants are language-generalists needs explanation (Chomsky's forte), but also the process by which they so quickly become culture-bound language-specialists, adopting a particular "native tongue" that permanently marks them. This indelible mark presents one of the deepest mysteries of early language development: try as one might, unlearning the accent or phonology of one's native tongue is virtually impossible. Henry Kissinger was not born with a German accent, nor Chomsky born with a Philadelphian one. These are not innate characteristics; once acquired, however, they have persisted over decades. Such is the mark of early learning.

The new data also suggest another shift from the standard nativist view. During early development, there is no compelling reason to postulate that the linguistic system functions independently of other cognitive and social systems. We will argue that although the language system may become modularized with development, infants do not begin life with a fully organized language module that is isolated from other aspects of cognition (Fodor 1983).

We will suggest a view that incorporates evolution, nativism and experience in the development of language. Our view embraces the notion that infants are born with abilities highly conducive to the development of language. We are nativists in this sense. These innate abilities *initially structure* the acquisition of language. However, infants' innate abilities do not *solely determine* language. Linguistic experience alters the system in profound ways. It fully restructures the system, and does so quickly, relatively permanently, and via an interesting mechanism that will be described here.

The theory and the arguments we present primarily address the phonetic level of language, the perception and production of the most basic units of language, the consonants and vowels of human speech. The phonetic level has advantages: one can study the comparative, developmental and cross-cultural aspects of the perception and production of speech. Even machines' capabilities to categorize the sounds of language can be tested. It therefore allows a comprehensive look at the underpinnings of humans' linguistic capacity. Our hope is that study of the phonetic level of language may inform theories of language acquisition at other levels.

The more specific goal is to elaborate further the *Native Language Magnet* theory of speech development first described by Kuhl (1992a, 1992b, 1993a, 1993b, 1994). A three-step process in the acquisition of speech is postulated: (a) innate perceptual boundaries exist that are tailor-made for language processing at the phonetic level; (b) exposure to ambient language results in stored representations that reflect the distributional properties of a particular language; and (c) the stored representations act recursively to alter the innately specified boundaries; they profoundly influence the subsequent perception and production of speech in relatively permanent ways. We believe that in early infancy, language acquisition is underpinned by a more general cognitive representational ability like the one described by Meltzoff (1990). This early representational system is polymodal – it is one to which all sensory modalities as well as the motor system has access. Moreover, the type of experience that influences speech representation entails a rather special interaction that occurs with conspecifics (Meltzoff and Gopnik 1993; Meltzoff and Moore 1995); a tape recorder presenting the sounds of language would not trigger it. The specific interweaving of what is "given by nature" and what is "gained from experience" is the story we will tell.

Nativism: Initial Structure for Phonetic Categorization

Infants have innate perceptual abilities that support the acquisition of language at the level of speech. Two pieces of evidence stand out: (a) categorical perception, a phenomenon showing that infants' perceptual systems partition sound to roughly define the phonetic categories of language; and (b) talker normalization, a phenomenon demonstrating that infants perceive their own vocalizations as "matching" adults' vocalizations, even though the two are physically very different. Even the most sophisticated computers have not succeeded in this special capacity for talker normalization. Yet human infants do so with ease. Such a biological endowment is necessary for infants to acquire the ability to speak themselves.

Categorical Perception

Tests of categorical perception (CP) use a continuum of speech sounds as stimuli. A series of sounds is generated by altering some acoustic variable in small steps. On one end of the series the sounds are identified as one syllable, the syllable /ba/ for example; on the other end of the continuum the sounds are identified as another sound, the syllable /pa/ (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967) (Figure 1).

Tests of CP ask listeners to identify each one of the sounds in the series. Early researchers expected that the sounds in the series would be perceived as changing gradually from /ba/ to /pa/, with many sounds in the middle of the series sounding "ambiguous" or a poor mixture of the two. That did not occur. Adults reported hearing a series of /ba/'s that abruptly changed to a series of /pa/'s. There was no in-between. When researchers asked listeners if they could hear the difference between two adjacent /ba/'s (or /pa/'s) in the series, they could *not* do so, even though the two /ba/'s (or /pa/'s) were physically different. Listeners did not hear differences between adjacent stimuli in the series until they heard a sudden shift – the change from /ba/ to /pa/. The fact that listeners' responses were "categorical" gave the phenomenon its name.

CP is sensitive to linguistic experience (Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura 1975). For adults, CP occurs only for sounds in their native language. When Japanese listeners were tested on a series of sounds that ranged from /ra/ to /la/ (a distinction that is not phonemic in Japanese), they did not hear a sudden change

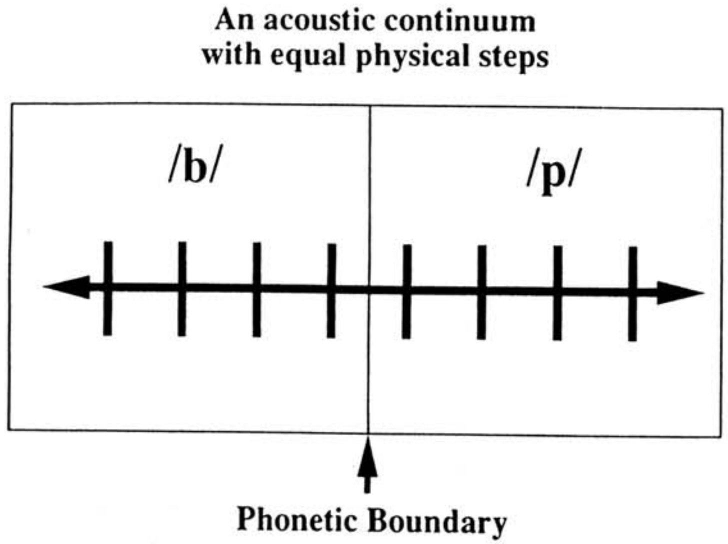

Figure 1: Categorical perception is tested using sounds from a computer-generated series. The sounds vary in equal steps along an acoustic dimension; however, perception changes abruptly at the location of the phonetic boundary between the two categories.

at the boundary between /ra/ and /la/. They heard no change at all. (This is why Japanese speakers tend to substitute /l/ for /r/ in speech.) Nonetheless, American listeners reported hearing a series of /ra/'s that changed suddenly to a series of /la/'s (Figure 2, top). The bottom half of Figure 2 compares the American and Japanese discrimination data. American listeners showed the characteristic peak in discrimination at the location of the /r-l/ boundary; Japanese listeners did not show this peak in discrimination at the phonetic boundary. Their performance in discriminating /ra/ from /la/ was at chance throughout the series (Figure 2, bottom).

The Americans were unlikely to have one set of innate endowments (a /ra/-/la/ detector) and Japanese another; that CP was language-specific suggested that it might be learned. Perhaps this learning arose as a result of hearing words with different referents contrasting /b/ and /p/ – like "bat" and "pat." If so, then very young infants would not be expected to show CP.

The relevant study was done by Eimas, Siqueland, Jusczyk and Vigorito (1971). Infants' responses to a /ba-pa/ series were monitored using a specially designed technique that relied on the measurement of sucking. The results showed that young infants demonstrated CP. Moreover, infants demonstrated the phenomenon not only for the

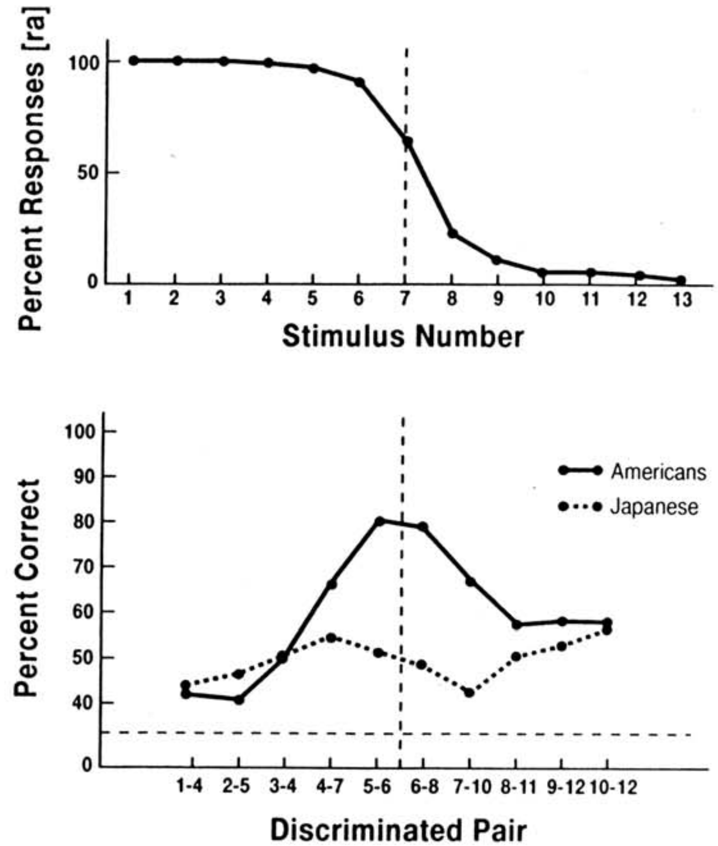

Figure 2: An example of categorical perception for the syllables /ra/ and /la/. American listeners identify sounds in the series (top) and show the characteristic peak in discrimination at the location of the phonetic boundary (bottom). Japanese listeners do not show the peak in discrimination. Redrawn, with permission, from K. Miyawaki, W. Strange, R. Verbrugge, A.M. Liberman, J.J. Jenkins and O. Fujimura (1975), An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–40.

sounds of their own native language, but also for sounds from many foreign languages (Streeter 1976, Lasky, Syrdal-Lasky and Klein 1975). Although adults were "culture bound," infants were not: they were primed to be members of any linguistic culture, "citizens of the world."

It can be concluded that infants' auditory perception is tailored to language processing at birth. Moreover, this does not depend on experience. Infants behave this way even for sounds they have never before heard. The puzzle that remains (to be discussed later) is when, why and by what mechanism adults "lose" a language-related ability that is present at birth.

Talker Normalization

The studies on CP show some rudimentary structure available to infants that helps them partition the perceptual space into gross divisions. However, perception of a phonetic category requires something more. In order to perceive a phonetic category, infants have to be able to perceive *similarity* among sounds that belong to a particular category, even though they are discriminably different. When different people produce the same vowel sound, one can hear the differences between them but one can also hear their identity. This is phonetic constancy despite auditory discriminability, categorization that renders discriminably different things equivalent.

This categorization ability is critical to infants' acquisition of speech. Infants' vocal tracts cannot produce the frequencies produced by an adult's vocal tract, so they cannot create the exact frequencies that an adult produces. Infants must hear the commonality between the vowels they are capable of producing and those produced by adults in order to learn to speak. Computers cannot yet be programmed to "perceive" these kinds of similarities across a wide range of talkers. Would naïve infants outperform the smartest computers in perceiving a perceptual similarity, *constancy*, for the same vowel produced by different talkers?

Kuhl demonstrated that infants have the ability to sort vowels by phonetic category regardless of the talker producing the sound (Kuhl 1979; 1991b). Figure 3 shows the results of two studies. Shown in the top panel are infant data from an /a-i/ categorization experiment (Kuhl 1979), and in the bottom panel, results from an /a-ae/ categorization experiment (Kuhl 1991b). In both, infants were initially trained to produce a head turn to a single vowel from Category 1 produced by a male speaker, but not to produce the head turn to a single vowel from Category 2 by the same speaker. The first and third panels show the results of the training data; infants master this task at the 90% correct level in short order. During the test phase of the experiment, novel exemplars are presented from both Category 1 and 2, produced by new male, female and child talkers. The results of these studies (second and fourth panels) demonstrated that infants generalize their head-turn response to the novel vowels of Category 1, but not Category 2, which is predicted by the hypothesis that infants are capable of perceptually sorting the novel vowels into two phonetic categories.

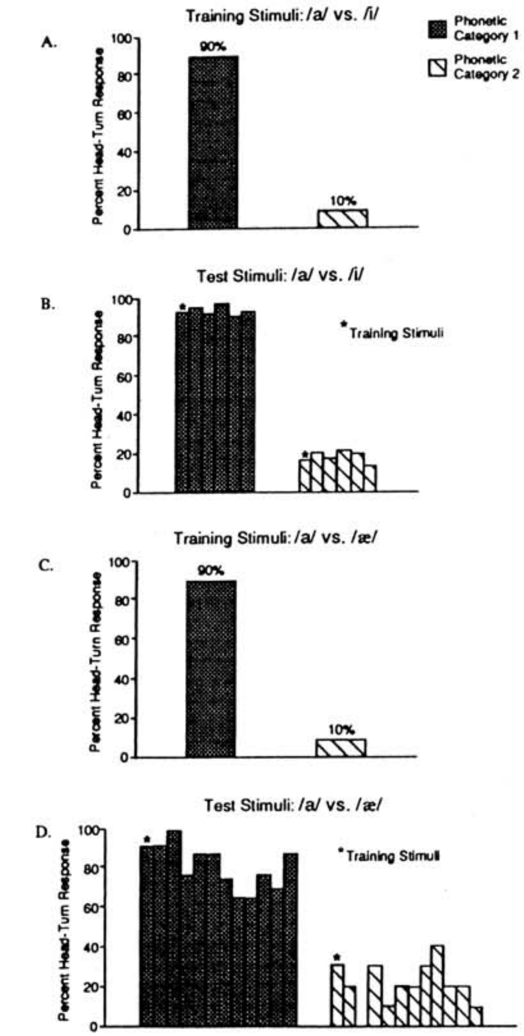

Figure 3: Categorization data from 6-month-old infants. Infants are trained until they reach 90% correct performance on the discrimination of two vowels spoken by a single speaker, panel A (/a/ vs. /i/), panel C (/a/ vs. /ae/). Infants were then tested using vowels produced by many speakers, including new male, female, and child speakers (panels B and D). Performance indicates that infants can perceptually sort vowels into phonetic categories regardless of the speaker who produces the vowel. From P.K. Kuhl (1991b), Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S.E. Brauth, W.S. Hall and R.J. Dooling (eds.), *Plasticity of Development*, 73–106. Cambridge, MA: MIT Press.