CHAPTER 16

INSIGHTS FROM ACQUISITION AND LEARNING

HOW PHONOLOGICAL REPRESENTATIONS DEVELOP DURING FIRST-LANGUAGE ACQUISITION

KATHERINE DEMUTH AND JAE YUNG SONG

SPEECH PROCESSING IN BILINGUAL AND MULTILINGUAL LISTENERS

PAOLA ESCUDERO

SECOND-LANGUAGE SPEECH LEARNING

RAJKA SMILJANIC

The contributions in this chapter discuss the role of language development in early acquisition, multilingualism, and second-language acquisition, and consider how these inform our understanding of core phonological questions. Together they paint a picture of the critical role of both production and perception in the learning of phonological systems and show how such acquisition studies provide insight into the nature of adult phonological structure.

16.1 How phonological representations develop during first-language acquisition*

Katherine Demuth and Jae Yung Song

16.1.1 Introduction

Little is known about the phonological representations that underlie children's early productions, and why variability in production persists even as the child progresses toward the mastery of language. Much of the early research on phonological acquisition focuses on children's production of segments, showing variability both between and within individuals (e.g. Smith 1973; Ferguson et al. 1992; Vihman 1993, 1996). Some of this research identified children's early motor-control limitations as a means for understanding their variable productions (e.g. MacNeilage 1980; Lindblom 1992). Others have shown that within-speaker variability is influenced by the frequency of lexical and syllable patterns in the ambient language (e.g. Beckman and Edwards 2000b; Levelt et al. 2000; Roark and Demuth 2000; Edwards et al. 2004; Storkel 2004; Zamuner et al. 2004; Munson et al., this volume). Still others have shown that the phonological contexts in which words and morphemes appear can have an enormous effect on whether a morpheme is apparently produced or not (e.g. Panagos et al. 1979; Bennett and Ingle 1984; Echols and Newport 1992; Gerken and McIntosh 1993; Rvachew and Andrews 2002). Variable processes of coda deletion and coda cluster reduction are also subject to contextual variation within certain dialects. This has been investigated for adults speaking American and British English (e.g. Roberts 1997; Foulkes et al. 2005; Docherty et al. 2006), and African-American English (e.g. Stockman and Vaughn-Cooke 1989; Wolfram 1991; Moran 1993; Bailey and Thomas 1998; Rickford 1999; Stockman 2006). Such adult

* We thank our collaborators Stefanie Shattuck-Hufnagel and Lucie Ménard for their contributions. This work was funded in part by NICHD grant #R01 HD057606.

variation is an important issue to keep in mind when examining child phonological and morphological development across dialects.

In this section, we argue that children's phonological representations as evidenced by their productions may be more intact than often assumed. We also suggest that conducting fine-grained acoustic analysis of child (and child-directed) speech holds the potential for better understanding children's developing phonological representations, and the factors that influence variability in production over time. We review below some of the traditional methods used, discussing some of their limitations, and then discuss recent laboratory phonology research examining the development of phonological representations as evidenced through production.

16.1.2 Traditional methods and some limitations

16.1.2.1 *Observational/longitudinal studies*

Many investigations of children's phonological development have been observational case studies, where longitudinal data is collected and developmental trends assessed. Some consist of diary studies (e.g. Deville 1891), whereas others consist of tape-recorded and transcribed child speech, using either orthographic (Brown 1973) or phonetic (IPA) form (Smith 1973). Such studies provide useful albeit impressionistic information about a child's language development, upon which many theoretical claims have been made. In fact, acoustic analysis is critical to fully analyze these data. For example, if the transcription indicates that the child produced no coda consonant on the word *dog*, it is impossible to know if the child's representation was really CV, or if there might have been vowel lengthening, indicating that the child has some knowledge of the "missing" coda consonant. Nonetheless, these types of spontaneous, longitudinal corpora can be extremely useful as pilot data for forming hypotheses about aspects of phonological development, which could be investigated under more controlled, experimental conditions. They are also useful in documenting individual differences in phonological development. However, for any specific research question there may not be enough tokens of the right type from spontaneous speech corpora to fully assess the extent of children's phonological knowledge.

16.1.2.2 *Experimental production studies*

Some of the concerns about sparse data can be addressed in cross-sectional experimental studies using elicited imitation or elicited production methods. This provides the opportunity for exploring children's phonological and morphological knowledge under controlled contexts at a given point in time. For example, Kirk and Demuth (2005) compared 2-year-olds' acquisition of segmentally similar consonant clusters at the beginnings and ends of words (*ski* vs. *ask*, *ax*). They found that children were better at producing consonant clusters word-finally, especially when these decreased in sonority (i.e. *ask* [æsk] was produced more accurately than *ax* [æks]). However, it is also possible that some of children's early cluster errors could be due to articulatory difficulty. For example, Kirk (2006) examined 2-year-olds' coda productions in monosyllabic and disyllabic non-words. They found better coda production in monosyllabic words, and in the final and stressed syllables of disyllabic words. Similarly, Song et al. (2009) found better production of third person singular *-s* in utterance-final compared with utterance-medial position—for both 1;10-year-olds and 2;3-year-olds (with the older children doing better overall). Taken together, these results suggest that children are more accurate in producing coda consonants in stressed and final syllables, which are longer in duration, thus allowing more time to complete the full articulation. On the other hand, many experimental production studies have not necessarily examined the data from a more fine-grained acoustic perspective.

16.1.2.3 *Experimental perception studies*

There is a wealth of literature examining the development of infants' perceptual abilities (see Holt, this volume; Munson et al., this volume; Maye, this volume). Some of this literature has focused on the development of native vs. non-native speech contrasts in infants under 1 year of age, showing that this can be influenced by segmental frequency effects (e.g. Anderson et al. 2003). It has been found that 19-month-olds have detailed subphonemic phonological representations that encode cues for place, manner, and voicing (White and Morgan 2008). However, mapping novel words onto objects appears to be challenging for 14-month-olds, indicating a heightened processing load that is only overcome around 20 months (Stager and Werker 1997; Swingley and Aslin 2000). There have also been several studies showing cross-linguistic differences in listening preferences for different types of lexical stress (see Nazzi et al. 2006 for review). However, there has been little investigation of infants' preference for other types of phonological units (though see Jusczyk et al. 2002).

In summary, much has been learned over the past twenty years about the course of phonological development. However, the methods used all exhibit certain limitations. The longitudinal studies have typically lacked an accompanying acoustic record and tend to involve small case studies. Likewise, many cross-sectional production studies have typically not exploited information in the acoustic signal when assessing children's phonological competence. They therefore miss potential covert contrasts the child may be making, presenting an incomplete and potentially misleading picture of what children know about phonological structure. They also tend to focus on one age, with little attention to development. Finally, although a few phonological issues have been examined in infant perception studies, these typically use non-linguistic measures, such as listening times or listening preference. Many of the experimental studies also report only group data, making it difficult to assess individual differences in phonological development. Nonetheless, our understanding of how and when children begin to develop grammatical competence at different levels of phonological structure is quickly evolving through the use of more widely available laboratory phonology methods, promising new and exciting results in the coming years.

16.1.3 Contributions from laboratory phonology

As outlined above, one of the challenges to understanding the development of phonological knowledge is that children sometimes make acoustically measurable distinctions corresponding to constrasts in adult speech but that are not perceived by the adult. This includes making subtle VOT distinctions for target voiced vs. voiceless onset stops, both of which tend to sound voiced to the adult ear (e.g. Macken and Barton 1980; Scobbie et al. 2000), and extrinsic vowel-duration distinctions before apparently missing voiced vs. voiceless codas (Weismer et al. 1981; Stoel-Gammon and Buder 1999). Young 1–2-year-olds have also been found to use spectral and durational cues to distinguish /gr/ from /gl/ in onset clusters (both heard as /gw/ by adults; Kornfeld 1971). Thus, children may acquire adult-like phonological contrasts earlier than often assumed, despite the fact that their early words often deviate from the adult form. Below we review further such evidence and discuss several possible factors that affect young children's production.

16.1.3.1 *The development of syllable and prosodic word structures*

Researchers have noted that children's early word shapes follow a systematic course of development. Drawing on data from English and Dutch (Fikkert 1994), Demuth (1995b) identified four stages in the development of words, suggesting that similar stages of development might be found in the acquisition of all languages. For example, Dutch-speaking children's early words expand from core syllables (CV) (e.g. [fa] for *olifant* 'elephant') to minimal words (bimoraic feet, i.e. CVC, CVV(C), or CVCV in Dutch) (e.g. [faut] 'elephant'), and eventually to larger, more complex phonological words (e.g. [olifant] 'elephant') as they progress in language acquisition.

While exploring four English-speaking children's attempts to produce coda consonants in monosyllabic CVC words such as *dog* [dɔg], Demuth et al. (2006) found that two of the children from 1–1;6 often lengthened the vowel when the coda is omitted, or added heavy aspiration or an epenthetic vowel to codas that were produced (e.g. CV ~ CVC ~ CVC^h ~ CVCV). Similar findings have been reported from corpus studies of other 1–2;6-year-olds (Vihman and Velleman 1989;

Goad and Brannen 2003). This raises the question of the nature of children's early syllabic representations, and whether these include coda consonants at all. Goad and Brannen (2003) proposed that heavy aspiration (typically appearing syllable-initially at this stage) occurring on the final consonant provides support that young children have only CV structure, and that apparent codas are actually onsets to an empty-headed syllable.

In contrast, Demuth et al. (2006) proposed that these children have a highly ranked NoCoda constraint, forcing output forms of CV, CVV, or CVCV. Given the high instance of vowel lengthening in the absence of a coda, they also suggested that English-speaking children may have an early awareness of word-minimality effects, where well-formed English lexical items must take the form of a bimoraic foot (Demuth 1995b; see Fikkert 1994 for similar explanations of early epenthesis in Dutch). Thus, children's early use of vowel lengthening and the addition of an epenthetic vowel (e.g. *dog* /dɔg/→ [dɔː] ~ [dɔgɔ]) could both be understood in terms of children's attempts to meet word minimality. Under all these approaches the assumption was that children had early limits on syllabic (and prosodic word) representations, and that this began to change around the age of 2–2;6, as more target-like coda consonants were produced. However, these studies did not conduct acoustic analysis to further investigate these issues.

Some of the first studies to explore these issues acoustically came from an investigation of word productions from 1–2-year-old Japanese-speaking children (Ota 1999). Since Japanese is a mora-timed language, the issue of when these children become sensitive to moraic structure was of interest. Using durational measurements, Ota (1999) found that Japanese learners showed moraic compensation when they omitted the coda consonant, lengthening the vowel, in effect to constitute two moras of structure. Specifically, he showed that the short vowel that preceded a missing coda nasal (CVØ) was significantly longer than a short vowel in an open syllable (CV) for all three children under investigation. Interestingly, such an asymmetry in vowel duration was not found when onset consonants were deleted, suggesting that the deletion of non-moraic segments does not lead to the compensatory lengthening of vowels. The findings suggest that Japanese children have an awareness of moraic representations or moraic weight of codas even when they cannot reliably produce the word-final consonants.

Similarly, Song and Demuth (2008) examined three English-speaking children's compensatory lengthening of vowels in the context of missing codas (e.g. *dog* /dɔg/ → [dɔː]). Languages like English require well-formed content words to contain a bimoraic foot with either a coda consonant (e.g. *tin* [tɪn]), or a tense (long or bimoraic) vowel (e.g. *tea* [ti]) (Hammond 1999). Thus, if lengthening selectively occurs with lax (monomoraic) vowels but not with tense (bimoraic) vowels, this would support the hypothesis that compensatory lengthening serves to preserve bimoraic or minimal word structure. However, if lengthening occurs across the board, this would indicate that increased vowel duration compensates for the omitted segment. The results showed that 1–2-year-olds lengthened both long and short vowels, suggesting that lengthening was compensating for the missing segment rather than the timing unit, i.e. mora (Stemberger 1992). This suggests that learning some of the language-specific constraints on prosodic word structure may take longer than previously assumed (Demuth 1995). However, it also provides support for the notion that these children have some representation for the missing coda.

16.1.3.2 *Limitations on the articulatory control of onset and coda consonants*

So far, we have provided evidence that children can exhibit adult-like representations of words even when their word production is not yet adult-like. This raises the question as to the nature of the factors that affect children's early productions. It is possible that there is a speech-planning explanation for these findings. That is, children might have a coda in their phonological representation, but not yet having the articulatory gestures needed to execute CVC especially within a multi-word utterance. The findings reported above, where morphemes were more accurately produced utterance-finally compared to utterance-medially, provide some support for this position. Furthermore, vowel epenthesis appears most often following voiced codas, and aspiration noise appears most after voiceless codas (Demuth et al. 2006), suggesting that processes of speakers using acoustic cue enhancement might be involved (Keyser and Stevens 2006). That is, the child may be trying to ensure that cues to the voicing of the final consonant are clearly perceived although the cues might not be quite adult-like.

Weismer et al. (1981) found that children who apparently "omit" word-final stops nevertheless produce a stop allophone in word-medial position (e.g. *do(g)* vs. *doggy*), indicating that /g/ must be part of the lexical representation of *dog*. This raises the possibility that some children's early attempted codas may include coda closure, but lack the acoustic cues expected by an adult. We are currently conducting acoustic and ultrasound analyses to see if there is evidence for an incomplete closure gesture at early points in development. If such covert contrasts were found, it would suggest that the acquisition of coda contrasts is a gradient rather than a discrete process, with "quasi codas" produced en route to full coda articulation (cf. Hewlett and Waters 2004). In addition, although most typically developing English-speaking children reliably produce coda releases by the age of 2, there is still some variability in the acoustic realization of coda stops. We are currently pursuing investigation of these issues in the acoustic record of both children and adults to better determine the development of acoustic cues to phonological contrasts, and the extent of individual variation (Demuth et al. 2009).

Further acoustic evidence of articulatory challenges faced by learners comes from Imbrie (2005), who compared ten children's variable productions of the onset stops /b, d, g, p, t, k/ at 2;6–3;6 using durational, amplitude, spectral, formant, and harmonic measurements. When these acoustic measures were interpreted in terms of the supraglottal, laryngeal, and respiratory actions that give rise to them, comparison with adult productions of the same words showed that children have acquired appropriate positioning of their primary articulators for producing a stop consonant. However, the children's gestures were still far from achieving the adult pattern even by the age of 3;6. For example, at this age children are still learning to adjust the tongue body during stop production, and the higher compliance of the articulators, smaller articulator size, and high subglottal pressure results in more tokens that have multiple release bursts and bursts that are shorter than those of the adult speakers. Longer VOT times and highly variable f0 suggest that children are still learning to adjust vocal fold stiffness and glottal spreading, as well as intraoral pressure. High variability in amplitude across an utterance suggests they are also still learning to control subglottal pressure. Thus, these children were less consistent than adults in controlling and coordinating certain aspects of their articulatory gestures, articulator stiffness, and respiration, though some aspects of the children's speech did become more adult-like over the course of the year of the study (see McGowan and Nittrouer 1988 and Nittrouer et al. 1989 for similar findings for fricatives).

Using the same methods, Shattuck-Hufnagel et al. (2011) examined children's coda consonant productions, focusing on cues to voicing distinctions. The findings indicate that 2;6–3;6-year-olds exhibit systematic acoustic cues to coda-voicing contrasts (e.g. *dog* vs. *duck*): an observable voice bar was more likely to precede voiced codas, whereas vowel glottalization was more likely to precede voiceless codas. Results from both 1;6–2;6-year-olds and their mothers' child-directed speech show similarities; the voice bar appeared more frequently before voiced compared to voiceless codas (Demuth et al. 2009). For mothers, the duration of the voice bar was also longer for voiced codas, and children showed a trend in this direction. However, only mothers showed a significantly higher use of vowel glottalization before voiceless codas. Thus, although younger children produce some acoustic cues to coda-voicing distinctions, other cues take more time to become adult-like.

These findings raise questions regarding the relationship between early articulatory gestures and phonological representations. Regarding tongue gestures, Gick (2007) examined ultrasound recordings of an 11-month-old child imitating productions of /r, l, w/. In accord with results discussed so far, he found that the child's production employed distinct articulatory traces and acoustic cues for each phoneme, despite the fact that the percept was not completely adult-like. On the other hand, Ménard et al. (2006) found that French-speaking 4-year-olds' CVC syllables were produced using different types of lip gestures than those of adults, and that children's stressed and unstressed syllables were less differentiated than for adults.

Preliminary study of two Canadian French children (aged 1;11 and 2;3) explored these issues in children's monosyllabic (CV, CVC) and disyllabic (CVCCVC) familiar words (Ménard and Demuth in preparation). The older child produced many word-medial codas, and had a distinct movement of the tongue for final VC as compared with final V. However, the younger child did not produce any codas in the disyllabic words he attempted. Furthermore, his vowels in the resultant CVCV productions were almost twice the duration of other vowels, showing compensatory lengthening. In addition, his tongue moved slightly toward the end of the vowel. This appears to be an articulatory gesture related to the attempted medial consonant, as confirmed by spectral analysis. These studies lay the groundwork for a more comprehensive investigation of young children's articulatory gestures using ultrasound. They also suggest the importance of conducting close acoustic and gestural analysis of apparently coda-less CVC utterances, for evidence of non-adult-like cue patterns and how they change as children master adult-like pronunciations.

16.1.3.3 *Context effects on the production and comprehension of grammatical morphemes*

Some of the variable production and comprehension of both phonological units and grammatical morphemes may be influenced by the prosodic context and/or utterance position in which they occur. Children have long been known to exhibit within-speaker variability in the production of English inflectional morphemes (e.g. Brown 1973). Many researchers suggest that this is due to incomplete semantic or syntactic representations. However, our recent study of third-person singular *-s* found that children were much less likely to produce this morpheme when it is a part of phonologically more complex codas (*hits* vs. *sees*), and in utterance-medial position as compared to utterance-final position (Song et al. 2009). This demonstrates that some of the within-speaker variability in the production of inflectional morphemes may be due to phonotactic complexity and positional effects. Hsieh et al. (1999) raise the possibility that this particular morpheme may be shorter in utterance-medial position. This could mean there is less time to produce it in utterance-medial position, resulting in more omission. Acoustic analysis of our stimuli used for both elicited production and comprehension experiments with 2-year-olds indicates that medial *-s* is indeed shorter than final *-s* (Song et al. 2009). This appears to have a negative effect on both production and comprehension of utterance-medial *-s* (Sundara et al. 2011). Interestingly, the effects of position are evidenced at the gestural level as well, in both older children (5–7-years-olds) and in adults (e.g. Nittrouer et al. 2005).

Gerken (1996) provided elicited production evidence for 2-year-olds showing earlier production of articles that are prosodically licensed as part of a disyllabic trochaic foot ($[$*hits the*$]_{Ft}$ $[$*piggy*$]_{Ft}$ vs. $[$*catches*$]_{Ft}$ *the* $[$*piggy*$]_{Ft}$). We examined longitudinal data to determine if 1–2-year-olds' use of articles would exhibit the same prosodic contextual effects in spontaneous speech. The results were confirmed for four of five children (Demuth and McCullough 2009). Interestingly, acoustic analysis of the productions from the fifth child showed a strong connection between prosodic organization and article production; her articles were produced as separate prosodic words at age 1;10, then became prosodified as part of a bimoraic foot (like the other children) at the age of 2. Little is known about young children's prosodic organization of grammatical morphemes, and how this develops over time. It is also unclear if children might go through a stage of development where they leave a "prosodic trace" for unrealized grammatical function items, such as that found in the omission of unfooted syllables for words like *(Lu)cinda* (Carter and Gerken 2004). Such a finding would provide additional evidence that children have some representation for the syllables and morphemes they omit. This is obviously a rich area for further research, using both longitudinal and cross-sectional methods.

16.1.4 Recent developments and future directions

16.1.4.1 *New methods*

New technological developments such as more accessible audio/video recording equipment and speech analysis software are beginning to address some of these limitations of previous longitudinal corpora. For example, the CHILDES database (MacWhinney 2000) now allows for both Unicode IPA transcription and the linking of audio/video files to the transcription record. This means that new databases, such as the Providence Corpus (English; Demuth et al. 2006) and the Lyon Corpus (French; Demuth and Tremblay 2008), are being donated with the audio files attached, allowing for a close examination of phonological and morphological development over time. This will permit much more extensive study of the acoustics of child and child-directed speech, and how this develops for the mothers and their children aged 1–3 years. The inclusion of the mother's speech in these corpora is particularly important, serving as a baseline for understanding the nature of the input.

Ultrasound methods are only now starting to be used to explore the nature of children's early phonological representations. With a small ultrasound probe placed under the chin, it is possible to collect both acoustic recordings of child speech and video recordings of tongue movements in a non-invasive manner. This can provide some idea of the types of articulatory gestures being made, and the extent to which these may be incomplete. For example, some children exhibit protracted problems with the production of glides, producing only some of the required articulatory gestures (Bernhardt et al. 2005). This method therefore holds the potential for providing a better understanding about the articulatory underpinnings of phonological development, and possibilities for remediation.

16.1.4.2 *Future directions*

To adequately address the nature of language acquisition it is critical to know more about the input children hear. Some suggest that child-directed speech is a form of "clear speech", with larger vowel space (Kuhl et al. 1997) and less reduction of segments (e.g. *want (h)im*) than those typically found in adult-directed speech (e.g. Bernstein-Ratner 1982, 1987). However, studies of Dutch child-directed speech suggest more reduction of vowels in grammatical morphemes (van de Weijer 1998). It would be therefore extremely helpful to know more about the acoustic/phonological properties of child-directed speech and the possible connections with individuals' phonological development, as well as how both change over time. Computational techniques are currently being developed that could eventually approximate an automatic alignment of phonemes with the acoustic signal, making it possible to examine a large amount of child-directed speech (e.g. Sjölander 2003). This in turn could shed light on the nature of the acoustic input language learners actually hear, providing a better understanding of how and when children develop adult-like phonological representations, both perceptually, and in production.

In summary, we have examined evidence from laboratory phonology research showing that children under the age of 3 have more robust phonological representations of syllable structures and words than evidenced from impressionistic studies of production. Examining the shape of children's early syllable and word productions is crucial to improving our understanding of the emergence of phonological representations. Several new data sources and methods are now making it possible to conduct laboratory phonology studies of phonological development in ways that were not possible before. This has brought with it an increasing number of studies from different languages, enriching our understanding of the acoustics of phonological development in a cross-linguistic context. The next decade promises to be an exciting one, with much more research on phonological development using laboratory phonology techniques. The results should provide a clearer picture of the course of phonological development at various levels of structure, and the implications this holds for later language development more generally.