5 Phonological acquisition

Mitsuhiko Ota

Introduction

This chapter reviews some important outcomes of developmental research on Japanese phonology, particularly in the area of early production. The review highlights the acquisition of several phonological phenomena in Japanese that are of particular interest from a crosslinguistic perspective, such as durational contrasts, mora-timing, and pitch accent. These are aspects in which Japanese differs typologically from English and several other European languages whose developmental data have been the empirical foundation of much theorizing in language acquisition. As such, the acquisition of these properties provides complementary evidence and insights that can further our general understanding of phonological development.

Segmental development

Development of the segmental inventory

Recent research on early speech production shows that language-specific segmental inventories begin to take shape during the babbling stage. Japanese is no exception. Babbling vocalizations in Japanese contain more nasals in comparison to those in English or Swedish, and fewer labials in comparison to those in French or English (Boysson-Bardies & Vihman, 1991). These findings indicate that children exposed to Japanese are tuning into the frequency distribution of the target segments even before they produce their first words.

Despite this precocious sensitivity to the language-specific phonemic inventory, children do not produce all segments or contrasts from the initial stages of word production. The order in which targetlike production of Japanese segments is achieved has been investigated in a number of cross-sectional studies, typically by means of a word elicitation task and accuracy measurement based on phonetic transcription (e.g. Murata, 1970; Nakajima et al., 1962; Owada, Nakanishi & Oshige, 1969; Sakauchi, 1967; Umebayashi & Takagi, 1965). Although there is some degree of variability across studies and individuals,

several general patterns are identifiable in the overall results. As a basis for comparison, the adult segmental inventory is given in (1).

(1) Phonetic inventory of (adult) Japanese

Vowels		Consonants				
i	ɯ	p b	t d		k g	
e	o	ɸ	s z	ʃ ʒ	ç	h
a			ts dz	tʃ dʒ		
		m	n		ŋ	
			ɾ			
		w		j		

Vowels are among the first sounds to be produced reliably. By the age of 2, the accuracy level of vowel production reaches 90 percent for most children. Of the consonants, those acquired early are the nasals ([m, n, ŋ]], stops ([p, b, t, d, k, g]), and alveolo-palatal affricates ([tʃ, dʒ]), while those acquired late are the three allophones of /h/ ([h, ç, ɸ]), the flap ([ɾ]), and alveolar sibilants ([s, z, ts, dz]). Misarticulation of these late-acquired segments is known to persist until around the age of 4 in some children. Several studies also report velar stops as appearing later than their alveolar and labial counterparts (e.g. Owada & Nakanishi, 1971; Ueda, 1996b).

One way to interpret the general acquisition order of Japanese sounds has been suggested by Ueda (1996a), using a feature-based model of segmental development. Under this view, inventories are expanded to include more phonemic contrasts by turning on the marked value of a feature. Thus, in Japanese the acquisition of fricatives ([+continuant]) implies the previous or simultaneous acquisition of stops ([−continuant]); the acquisition of velars ([dorsal]), that of alveolars ([coronal]); and the acquisition of sibilants [+strident], that of nonsibilants [−strident]. This model also correctly predicts that, before a target segment with a marked feature value becomes available, it is likely to be replaced by a segment with the unmarked (default) value of the same feature. Thus, a frequent substitute for a sibilant fricative ([+continuant]) is a stop ([−continuant]) (except before, [i] where stops are affricated), as shown below.

(2) Substitution errors: s, ʃ → t, tʃ (Ito, 1990; Owada, Nakanishi & Oshige, 1969)

a. sakana → [takana]	"fish"
b. kiʃa → [kita]	"train"
c. hoʃii → [hotʃii]	"I want it"

Similarly, the most common substitute for the flap [ɾ] ([+sonorant]) is [d] ([−sonorant]) (Murata, 1970; Sakauchi, 1967; Ueda, Ito & Shirahata, 1998; Umebayashi & Takagi, 1965). This substitution, known in the literature as

"rhotacism," has two patterns. In the first type, all instances of target /ɾ/, regardless of their position, are realized as [d]. In the second type, as illustrated in (3), the two sounds stand in complementary distribution whereby [d] occurs word-initially, and [ɾ] word-medially.

(3) A case of rhotacism (Ueda, Ito & Shirahata, 1998)

a. rappa → [dappa]	"trumpet"
b. denʃa → [denʃa]	"tram car"
c. terebi → [terebi]	"TV"
d. budoo → [bɯroo]	"grape"

Ueda and Davis (1999, 2001) present an analysis of rhotacism within the framework of Optimality Theory, and demonstrate that the substitution can be modeled in terms of the interaction of two constraints: one that militates against flaps and one that prefers sonorant segments in intervocalic positions. Different rankings of these constraints give rise to the two attested patterns of rhotacism. The key insight in this analysis is that a phonological process that is seemingly specific to child Japanese may be a product of universal but conflicting tendencies in human languages. In this respect, it is interesting to take note of the similarities between the distribution of [ɾ] in (3) and the environment where flapping occurs in adult American English.

Development of durational contrasts

Japanese has durational contrasts in both vowels and consonants, as illustrated by lexical pairs such as /to/ "door" vs. /too/ "ten," /saka/ "slope" vs. /sakka/ "writer," and /ama/ "nun" vs. /amma/ "masseur." Studies show that the distinction between short and long segments emerges early in children's word production but the actual phonetic values of the contrasts take several years to converge on those of adults. For instance, Ota (2003a) found that the duration of the syllables produced by 1-year-olds for CVV targets was already longer than that for CV targets, but the ratio between the two was 1.3:1 to 1.8:1 – a difference that falls short of the average adult ratio, which ranges from 1.7:1 to 2.1:1 (Warner & Arai, 2001). Similarly, Aoyama (2000) showed that the ratio of geminate ([nn]) to singleton ([n]) nasal durations was 1.4:1 at age 3, and although the ratio increased to 1.6:1 at age 5, it was still lower than the value recorded by the adult controls (2:1). Therefore, the acquisition of durational contrasts in Japanese follows the developmental pattern attested in many other segmental contrasts, including the voice/voiceless distinction in English (Macken & Barton, 1980), where the phonetic realizations of a phonological contrast show a gradual shift toward the target values.

Prosodic development

Prosodic organization of early words

A widely discussed issue in recent phonological acquisition research is the extent to which the structure of early words is organized in terms of the same prosodic units found in adult phonology (Demuth, 1995; Fikkert, 1994; Kehoe & Stoel-Gammon, 1997). These units include the prosodic word, the foot, the syllable and the mora, which are organized in a hierarchical structure. In addressing this issue, several studies have examined the effects of the so-called "minimal word" condition (Kehoe & Stoel-Gammon, 2001; Salidis & Johnson, 1997). If all the prosodic units mentioned above are part of the child's phonological system and realized initially in their unmarked forms, we predict that early words must be at least two morae long, the minimal size of a prosodic word that contains a binary foot.

Testing this prediction in child English has proven to be problematic because the phonetic evidence for morae in child English is not clear and no adult English lexical words are smaller than two morae. Data from Japanese, on the other hand, are ideal for this purpose. Based on the pattern in adult Japanese, we can take a (C)V syllable to be monomoraic, and all other syllable types, i.e. (C)VV, and (C)VC, to be bimoraic. In addition, there are monomoraic lexical words in the language (e.g. /me/ "eye," /te/ "hand," /ki/ "tree") which should undergo some prosodic modification if early words obey the bimoraic minimality condition. To this end, two crucial observations have been made. First, even though children frequently truncate words in their production, they rarely reduce them to a size smaller than two morae (Ito, 2000; Ota, 2003a). Second, and more importantly, some 1-year-olds lengthen the vowel in monomoraic lexical words, e.g. /me/ → [me:], /te/ → [te:] (Kawakami & Ito, 1999; Ota, 2003a). These lengthened CV target words also exhibit the type of falling pitch contour carried only by words that are at least two morae long. Taken together, these observations indicate that the production of 1-year-old Japanese children is subject to the bimoraic minimal-word constraint, lending support to the view that early words are already organized in terms of adultlike prosodic units.

Another characteristic of the shape of early words in Japanese is the prevalence of heavy–light (or bimoraic–monomoraic) and heavy–heavy (or bimoraic–bimoraic) syllable sequences. Indeed a large number of words unique to young children's lexicon ("baby-talk words") follow this pattern, e.g. /mamma/ "food," /ombu/ "carry-me," /okki/ "getting up," /sikko/ "wee-wee," /pompon/ "tummy," /oppai/ "breast." But a tendency toward heavy–light (HL) and heavy–heavy (HH) forms is also found in truncation (e.g. /tamago/ → [ta:ma] "egg") and insertion errors (e.g. /kaban/ → [kamban] "bag") (Ota, 1998, 2003a). Furthermore, there is experimental evidence that at 8–10 months, children show perceptual preference for HL over other trimoraic forms such as LH and LLL (Hayashi, Tamekawa & Mazuka, 2000). Interestingly, the frequencies of HL and HH words are highly inflated in caregiver speech. While 1.2% of trimoraic lexical items are HL and 0.5% of quadrimoraic lexical items are HH in the adult Japanese lexicon, these proportions jump to 80% and 60%, respectively, when type counts of caregiver vocabulary are examined (Hayashi, Tamekawa & Mazuka, 2000). Therefore, the status of HL/HH patterns in early Japanese words raises a chicken-and-egg question about the relationship between the prosodic shape of early words and the linguistic input from the adult: do children favor certain word forms because many salient lexical items used by caregivers are of those structures, or do adults create baby-talk words that reflect children's preferences for particular prosodic shapes?

Development of speech segmentation

Speech errors, stuttering, blending, and other psycholinguistic evidence show that adult speakers of Tokyo Japanese tend to segment words at points that correspond to mora boundaries, often breaking the integrity of the syllable, e.g. *to-o* "ten", *ho-n* "book" (Kubozono, 1989a, 1995, this volume, chapter 26; Warner & Arai, 2001). How this processing pattern develops has been a focus of debate. A potential source of influence is orthography (Beckman, 1995). The two *kana* syllabary systems in Japanese assign separate symbols to all moraic units (i.e. any CV unit, the second half of a long vowel or a diphthong, coda nasal, and the first half of a geminate). It is possible that these writing systems bias Japanese speakers toward a mora-based segmentation pattern. However, recent research indicates that the mora-based segmentation pattern appears in some Japanese-speaking children even before they are able to read (Ito & Kagawa, 2001; Ito & Tatsumi, 1997; Kawakami & Ito, 1999). In one study (Ito & Tatsumi, 1997), the children (age 3–5) were trained to segment words consisting only of monomoraic syllables (e.g. /hasami/ "scissors" → *ha-sa-mi*), and then asked to apply the procedure to words containing bimoraic syllables (e.g. /suika/ "watermelon," /ringo/ "apple," /boosi/ "hat," and /happa/ "leaf"). Although only eight of the twenty 4-year-old subjects could read *kana*, all of them separated the bimoraic syllables after the initial CV (i.e. *su-i-ka*, *ri-n-go*, *bo-o-si*), except for syllables closed by a geminate.

It appears, then, that preliterate 4-year-old children already have some awareness of mora boundaries; but where does this awareness originate? One explanation is that the high frequency of CV syllables in Japanese predisposes the learner to parse speech into units of CVs (Kubozono, 1995). Alternatively, the segmentation pattern may arise from the temporal organization of the language, which tends to be synchronous to mora units (Otake et al., 1993). Yet another possible source is the presence of phonological and phonetic

phenomena in Japanese that apply to segments within a mora but not across two morae (Kubozono, 1999, this volume, chapter 26). One or more of these factors may lead the learner to tune in to the moraic segmentation of speech in Japanese.

Development of pitch phonology

Pitch patterns in Japanese are determined by a number of factors including lexically specified pitch accents, tonal markers of phrasal boundaries, and the morphological composition of words. The contour pattern of children's early utterances reveals the process by which children unravel these different aspects of pitch phonology.

Isolated monomorphemic words in Tokyo Japanese can be marked by a lexical pitch accent, which is realized as a falling contour, and a lowering of (short) syllables at the beginning of a prosodic phrase. The interaction of these features is illustrated in (4), where the position of the lexical pitch accent is indicated by a diacritic.

(4) Pitch contours of isolated words in Tokyo Japanese

a. áme "rain" (initial accent)

b. mamé "bean" (final accent)

c. ame "candy" (unaccented)

In spontaneous speech produced by 1-year-olds, most initially accented disyllables (see (4a)) are produced with an adultlike falling contour (Hallé, Boysson-Bardies & Vihman, 1991; Ota, 2003b). However, the contours produced for final-accent and unaccented disyllabic words tend to have a flat profile, even though targetlike productions should show a rise into the second syllable (see (4b) and (4c)). This suggests that although most 1-year-old Tokyo Japanese-speaking children have acquired the lexical pitch accent, they are still in the course of developing adultlike realization of the phrasal aspects of pitch phonology. The late development of the phrase-initial lowering may be due to its variant realization (i.e. it is blocked by an initial accent) or its lower salience (i.e. the change in fundamental frequency of the rise is smaller than that of the lexical pitch fall).

Much less is known about how children acquire the pitch patterns of morphologically complex words, but crossdialectal research on the development of compound accent rules indicates that the process involves several mechanisms. Experimental studies show that 3- to 5-year-old Kagoshima children produce

compounds with adultlike accent assignment, but Tokyo and Kyoto children of the same age range have not fully mastered their compound accent rules (Shirose & Kiritani, 2001; Shirose, Kakehi & Kiritani, 2001). This is likely to reflect the different degrees of complexity and predictability of compound accent in these dialects. Interestingly, in both Tokyo and Kyoto Japanese, nonadultlike accent assignments tend to fall on the antepenultimate mora. Shirose, Kakehi, and Kiritani (2001) suggest that this pattern emerges from a default accent assigned to the head of a nonfinal foot, a position that is most unmarked with respect to several universal constraints on metrical structure. Thus, the acquisition of compound accent rules in Japanese offers a useful testing ground for the interaction of language-specific properties and universal tendencies in prosodic acquisition.

Conclusions

Research on the acquisition of Japanese phonology has generated a considerable amount of empirical findings with important implications for general models and theories of phonological acquisition. Still, much ground in the field is left for further investigation. In particular, the development of morphologically conditioned phonological processes, such as rendaku (sequential voicing), truncation, reduplication and inflectional morphophonology, is a domain that calls for more attention. As discussed in the context of pitch phonology acquisition, crossdialectal comparison is a great source of hypothesis-testing, but remains largely untapped. Further research in these areas will most certainly shed new light on the general mechanisms that underlie the development of phonology.