2 Phonological Acquisition

Phonology concerns the regularities and rules governing pronunciation of words, phrases and sentences. This chapter concerns the development of phonological knowledge. Evidence from very early perception of sound units and spontaneous babbling of speech sounds provides support for the view that early perception and production must be characterized in terms of distinctions that have a prominent place in adult phonological systems. Later phonological development is not well understood, but none the less reveals on occasion striking abilities to abstract rules from the data, within the bounds of conformity to the general dictates of phonological theory.

2.1 Speech Sounds

Words can be analysed as a sequence of perceptually discrete units – for example the word ‘cat’ can be analysed as a sequence of three distinct segments. This section describes some of the major properties of sound segments, properties essential to understanding how segments function in phonology. The description is based on sound segments found in English; language variation is taken up later.

2.1.1 How Sounds are Made

The vast majority of speech sounds, and all English sounds, involve pushing air out from the lungs and through the mouth (or nose and mouth). The quality of the sound will depend on the shape of the resonance chambers (the mouth, pharynx and nose; see figure 2.1) and on whether or not the airstream is obstructed.

The most basic distinction is between vocalic and consonantal sounds. In the articulation of vowels, the airstream is interfered with only at the glottis (see figure 2.1); two bands of ligament and elastic tissue, called the vocal cords, obstruct the larynx and the airstream must force its way between them when a vowel sound is articulated, causing them to vibrate. But there is no
Figure 2.1  The human vocal tract

Language Files, second edition, Ohio State University

further obstruction as air passes out through the mouth. By contrast, for most consonants, there may or may not be obstruction at the glottis, but there will be some obstruction of the air passage through the mouth or pharynx.

Vowel sounds  The quality of an individual vowel sound depends on the position of the body of the tongue – whether it is projected towards the front of the mouth or bunched towards the back, and whether it is held relatively high or relatively low. English has a series of front vowels and a series of back vowels. Table 2.1 lists the main English vowels and gives examples of English words in which they occur. (The phonetic symbols in table 2.1 and below are those of the International Phonetic Alphabet; they provide a 'one sound, one symbol' system that eliminates the inconsistencies of English spelling, where in many cases the same letter is used for different sounds and different letters.
Table 2.1 English vowels

<table>
<thead>
<tr>
<th>Tongue height</th>
<th>Tongue projection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Front</td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>[i] beet</td>
</tr>
<tr>
<td>l</td>
<td>[l] bêt</td>
</tr>
<tr>
<td>Mid</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>[e] bête</td>
</tr>
<tr>
<td>i</td>
<td>[i] bet</td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

*The examples given are based on standard southern British pronunciation. Details and variants of pronunciation in different varieties of English should not affect the points that follow.

* The vowels [æ], [a] are in fact diphthongs: combinations of a vowel with a glide sound (see the following section for a brief description of glides); so too are [i] and [u] though the glide is brief.

* The symbol [a] represents a more central version of [o] and occurs as the first element in the diphthong in, for example, the word ‘pie’. It is often used to represent [o].

are used for the same sound.) The reader can check that the tongue height positions, etc. are as they are given in table 2.1 by saying these words and comparing the position of her tongue for the different sounds. The mid-vowel [a] (the first and third vowel in the word ‘banana’) is used in the normal pronunciation of many vowels when they do not bear stress. All the back vowels in English (except [a]) are pronounced with lip-rounding, and the front vowels without lip-rounding. This type of asymmetry is normal in languages, although front round vowels are not uncommon.

Consonants The quality of a consonant sound will depend on the type of obstruction and the place at which the obstruction takes place. Stop sounds involve very brief complete blockage of air; fricative sounds involve a loose occlusion rather than an absolute blockage. In English, stops are formed by closing the two lips together ([p], [b]) or by touching the tongue to the roof of the mouth at the alveolar ridge ([t], [d]) or the velum ([k], [g]). In English, fricatives are formed by loose contact between the upper teeth and lower lip ([f], [v]), between the tongue and upper teeth (the initial sounds in ‘thin’ and ‘then’, [θ] and [ð]), between the tongue and alveolar ridge ([s], [z]), and between the tongue and a position slightly to the back of the alveolar
ridge (palato-alveolar fricatives, such as the first sound in 'sure' /ʃ/ and the medial sound in 'measure' /ʒ/).

All of the above examples of stops and fricatives come in pairs – [t]/[d], [s]/[z], etc. The difference between the sounds in each pair is not a matter of place of obstruction in the vocal tract or degree of obstruction, but of whether or not the passage of air is additionally interfered with at the glottis. If the vocal cords are close enough together to be set in motion, as they are in the articulation of vowels, then as the air moves through the narrow gap between the cords, they will vibrate and the sound will be voiced. If the vocal cords are spread apart, allowing air to pass without obstruction at the glottis, the sound will be voiceless. In the pairs of stops and fricatives given above, the first member is voiceless and the second voiced; thus [t] is a voiceless alveolar stop and [d] is its voiced equivalent.

In nasal consonants there is free passage of air through the nose. For all the sounds described above, the soft part of the back region of the roof of the mouth (the velum) is raised up, so that the entrance to the nasal cavity is blocked and air cannot escape through the nose. If the velum is lowered, then air can pass through the nose and the result is a nasal sound. Each of the stops in English has a nasal variant: [n], the first segment of ‘night’, is the result of a stop articulation at the alveolar ridge plus free passage of air through the nose; [m], the initial sound in ‘might’, and [ŋ], the last segment in ‘tang’ (written ‘ng’), are nasals resulting from labial and velar stop articulations, respectively. These nasals are all voiced, as is usually (though not invariably) the case in languages of the world.

Other types of consonantal sounds include affricates, where there is complete closure followed by a gradual, fricative release. English has the palato-alveolar affricates exemplified by the first sounds in ‘church’ ([ʃ], written ‘ch’) and ‘judge’ ([ʤ]). There are also liquid sounds such as English [l] and [r], formed with semi-free passage of air and somewhat similar acoustically to vowels. Glide sounds are also more similar to vowels than other consonants; like vowels they are made without obstruction in the vocal tract. The glides [i] and [w] such as the initial sounds in English ‘you’ and ‘witch’ are formed by raising the tongue towards the front and back of the mouth respectively and moving rapidly to or from the position of the following or preceding vowel: hence the term glide. The back glide [w] is lip-rounded.

Table 2.2 groups the English consonantal sounds described above by place of articulation and manner of obstruction.

2.1.2 Phonetic Features

A sound segment can be represented as a cluster of properties – a set of plus
Table 2.2 English consonants: place and manner of articulation\(^a\)^b

<table>
<thead>
<tr>
<th>Manner</th>
<th>Labial</th>
<th>Labio-dental</th>
<th>Interdental</th>
<th>Alveolar</th>
<th>Palato-Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p, b</td>
<td>f, v</td>
<td>e, δ</td>
<td>s, z</td>
<td>f, 3</td>
<td>k, g</td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquid</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The voiceless member of voiceless–voiced pairs is listed first in the table.

\(^b\) Equivalent symbols commonly used in North American transcription are: $\odot = f$; $\odot = s$; $\odot = f$; $\odot = d$; $\odot = j$.

or minus specifications for features that refer to the articulatory or acoustic quality of the sound segment, or its ‘function’ in a syllable.

The basic distinction between consonant and vowel sounds is represented in terms of the feature $[\pm$ consonantal]. If we add a second feature $[\pm$ syllabic], reflecting roughly how central in a syllable an element is, we can make a three-way distinction between vowels, consonants, and glides, as shown in (1).

\[
(1) \quad \text{vowels} \quad \text{consonants} \quad \text{glides} \\
\begin{array}{ccc}
-\text{consonantal} & +\text{consonantal} & -\text{consonantal} \\
+\text{syllabic} & -\text{syllabic} & -\text{syllabic}
\end{array}
\]

Individual segments will be specified for values of additional features, sufficient to characterize each sound uniquely. The full set of features needed to pick out each separate sound in the repertoire of English sounds need not be listed here (and the exact membership of the set is a matter of debate). The features in (2) will be sufficient with respect to the discussion in the following sections:

\[
(2) \text{ for vowel sounds} \\
\begin{align*}
[\pm \text{high}] & \text{ high vowels are } [+\text{high}]; \text{ mid- and low vowels are } [-\text{high}];
\end{align*}
\]
low vowels are \([\pm\text{low}]\); mid- and high vowels are \([-\text{low}])\; \text{and}\; \text{back}\; \text{and}\; \text{central} \text{vowels are} \([\pm\text{back}]); \text{front} \text{vowels are} \([-\text{back}].

\text{for consonantal sounds}

\([\pm\text{continuant}]\) \text{stops are} \([-\text{continuant}]); \text{fricatives are} \([\pm\text{continuant}]\); \text{labial} \text{and} \text{alveolar} \text{sounds (pronounced in} \text{the front of} \text{the mouth)} \text{are} \([\pm\text{anterior}]); \text{palato-alveolar}, \text{palatal} \text{and} \text{velar sounds are} \([-\text{anterior}].

\([\pm\text{coronal}]\) \text{alveolar} \text{and} \text{palatal} \text{sounds are} \([\pm\text{coronal}]) \text{(produced with} \text{obstruction by the blade/tip of} \text{the tongue)} \text{and} \text{labial} \text{and} \text{velar} \text{sounds are} \([-\text{coronal}];

\([\pm\text{voice}]\) \text{[+ voice] sounds are produced with} \text{vibration} \text{of} \text{the vocal cords, as} \text{described above;}

\([\pm\text{nasal}]\) \text{[+ nasal] sounds are produced with} \text{passage} \text{of} \text{air through the nasal cavity, as} \text{described above.}

In terms of these features, the word ‘cat’ can be represented as a sequence of segments, each segment characterized by a set of feature specifications that distinguishes that sound from other sounds used in English:

\begin{equation}
\begin{bmatrix}
\text{\([+\text{consonantal}]\)} & \text{\([-\text{consonantal}])}\ & \text{\([+\text{consonantal}]\)} \\
\text{\([-\text{syllabic}]\)} & \text{\([+\text{syllabic}]\)} & \text{\([-\text{syllabic}]\)} \\
\text{\([-\text{anterior}]\)} & \text{\([-\text{high}]\)} & \text{\([+\text{anterior}]\)} \\
\text{\([-\text{coronal}]\)} & \text{\([+\text{low}]\)} & \text{\([+\text{coronal}]\)} \\
\text{\([-\text{nasal}]\)} & \text{\([-\text{back}]\)} & \text{\([-\text{nasal}]\)} \\
\text{\([-\text{voice}]\)} & \text{\([-\text{voice}]\)} & \text{\([-\text{voice}]\)}
\end{bmatrix}
\end{equation}

2.2 Phonetics, Phonology and Language Variation

Phonetics is concerned with the characterization of speech sounds — how they are produced and perceived and what their acoustic properties are; phonology is concerned with how sounds are used to distinguish meaning and with the rules governing the distribution of sounds and the properties of segments and strings of segments in languages.
Those phonetic properties of words that cannot be predicted by rule must be entered in the lexicon (dictionary) of a language; a representation such as that given above for the word ‘cat’ must be part of the dictionary entry of that word. Phonetic features that are specified in the dictionary entries for a particular language include only the distinctive features of the language – features that do service to distinguish between different words or morphemes. For example, *voice* is one of the distinctive features in English; if we change the feature specification for voicing in the last segment in ‘cat’, a different sound segment results, and the difference in sound corresponds to a difference in meaning (‘cad’ does not mean the same thing as ‘cat’):

\[
\begin{array}{c|c}
\text{t} & \text{d} \\
\hline
\begin{array}{c}
+ \text{consonantal} \\
- \text{vocalic} \\
+ \text{anterior} \\
+ \text{coronal} \\
- \text{nasal} \\
- \text{voice}
\end{array} & \begin{array}{c}
+ \text{consonantal} \\
- \text{vocalic} \\
+ \text{anterior} \\
+ \text{coronal} \\
- \text{nasal} \\
+ \text{voice}
\end{array}
\end{array}
\]

Other, predictable aspects of pronunciation will not be included in the dictionary representation, but are provided for by phonological rule. Thus there is a difference in the pronunciation of voiceless stops in English in initial position and after a consonant. Initial stops in English are *aspirated* – pronounced with a puff of air on the release of the stop; stops occurring in non-initial position (after the sound [s]) are unaspirated. This rule-governed difference in pronunciation of, for example, [k] in ‘cat’ versus [k] in ‘scatter’ or ‘skittles’ can be accounted for by a phonological rule that adds the feature [+ aspirate] in the appropriate environment. The occurrence of aspiration is predictable and non-distinctive in English: if a word such as ‘cat’ is pronounced without aspiration on the initial [k], then we still perceive the speaker’s utterance as an utterance of the word ‘cat’, although with deviant articulation.

In sum, the sound segments that compose words as they are uttered in a language can be represented in terms of phonetic features, some of which will be part of the word’s dictionary entry and others of which will be specified by phonological rule.

Languages vary with respect to the use to which they put phonetic features. Some features and feature combinations may not be used in a language. Thus English has no velar fricatives, although velar fricatives are not uncommon in languages. (The voiceless velar fricative does occur in dialects of English, as in Scots pronunciation of the last segment in ‘loch.’) Features may also be
distinctive in one language and non-distinctive in another. The feature
[± aspirate] is an example; although in English the occurrence of aspiration
is non-distinctive and predictable by rule, in other languages aspiration is used
distinctively and the difference between, for example, [k] and [kʰ] (aspirate
[k]) can signal a difference in the meaning of words (Thai and Sesotho, a
language spoken in southern Africa, are examples of languages that use aspira-
tion distinctively.) Finally, languages may differ with respect to the exact
articulatory and acoustic values that are assigned to plus and minus values of
phonetic features, a fact that is related in part to choice of which features are
distinctive. In the next section we will show examples of this type of variation
and infants' handling of the pertinent contrasts.

2.3 Categorical Perception in Adults and Infants

Phonetic feature specifications as we have considered them thus far are binary
– plus or minus. There is evidence that the human speech-perception device
imposes precisely such a type of yes/no distinction, dividing up speech sounds
into categories that reflect a qualitatively arbitrary distinction as to where the
boundary between one type of sound and another is drawn.

The acoustic cues for speech sounds are complex and may vary according
to the position of the sound in the syllable and the nature of the sounds
adjacent to it. Here we will consider only the perception of the
voiceless–voiced distinction in syllable-initial position; this contrast offers
some of the clearest examples of the phenomenon of categorical speech
perception.

Perception of voicing is dependent on the timing relation between the
release of the stop closure and the onset of voicing.¹ Experiments using syn-
thetic speech stimuli have shown that for English speakers a stop sound in a
stop-plus-vowel sequence will be perceived as voiced if voicing begins within
about 30 milliseconds of the release of the stop closure; if the voicing begins
more than 30 milliseconds after the release of the stop closure, the stop will
be perceived as voiceless (see Lisker and Abramson 1970). The change in
perception is quite abrupt. A difference of 10 milliseconds of voice onset time
(VOT) in the critical 30 millisecond VOT region will produce a dramatic
change in the perception of sounds as voiced or voiceless; a change of similar
magnitude on either side of the critical region will not produce such a change
in voicing perception. Perception of voicing is thus categorical: we divide
speech events up into discrete categories, such as voiced and voiceless, based
on a sharply defined point along the relevant acoustic parameter.
Eimas et al. (1971) performed a clever experiment, demonstrating that infants of one to four months are sensitive to the boundary that governs adult perception of voicing in English. Infants can be motivated to suck on a pacifier by an auditory feedback; when the infant sucks with sufficient force, he hears a sound. Typically, when an infant catches on to the relation between sucking and feedback, there is an initial period in which sucking rate increases. Following this, rate of sucking declines, presumably because the child becomes habituated. In Eimas et al.'s study, the pattern of increased sucking rate followed by decline was established for each subject with one stimulus type. When the sucking rate for the first stimulus type had declined by 20 per cent or more for two minutes compared with the minute preceding, a second stimulus was then presented for four minutes. Increase in the sucking rate at the point of changeover can be interpreted as evidence that the infants perceived the difference in the stimuli.

The stimuli in Eimas et al.'s study were synthetically produced syllables consisting of a labial stop plus a low back vowel. Six different stimuli were produced by varying the VOT. Voicing began 20 milliseconds before the stop release, at the stop release, and 20, 40, 60 and 80 milliseconds after the stop release. For English-speaking adults, stimuli with a VOT of up to and including 20 milliseconds after the release will be perceived as [b]; stimuli with a VOT of 40 milliseconds or greater will be perceived as [p]. There were two experimental groups of infants; the first group received stimuli with a VOT of 20 milliseconds after the stop release, followed by stimuli with a VOT of 40 milliseconds after the stop release. The two sets of stimuli for this group thus straddled the boundary for adult perception of [p] versus [b]. The second experimental group received two sets of stimuli that did not straddle the adult boundary — either stimuli with a VOT of 20 milliseconds preceding the release and stimuli with simultaneous voicing and release (both [b] to the adult ear) or stimuli with VOTs of 60 and 80 milliseconds after the release (both [p] to adults). A third, control, subject group heard the same stimulus throughout (different children hearing each of the six different stimuli used for the experimental groups).

Figure 2.2 displays the mean change in response rate for the two-minute intervals before and after the change in stimuli (or the point at which the change would have occurred, for the control group). The figure plainly shows that a change in stimuli that crossed the adult boundary for [b] vs. [p] produced a marked increase in sucking rate; a change that did not cross the adult boundary produced no such increase (the small rise for one-month-olds was non-significant). The control group showed no increase, but a continued decline in sucking rate, as would be expected for the continued presentation of the same stimulus.

Eimas et al.'s study involved infants whose exposure to their native
language was minimal and who had not yet begun to produce speech-like sounds. Thus, soon after birth we discriminate between sounds in a way that reflects the type of distinction employed in adult phonological systems and encoded in linguistic representations by means of binary (plus/minus) phonetic features.

Many studies have followed on Eimas et al.’s initial findings, using a variety of experimental techniques (see Eimas 1985 for an overview). Categorical discrimination of sounds has been shown in animals (chinchillas) as well as human infants (Kuhl and Miller 1975), suggesting that the mapping between linguistic distinctions and phonetic values may draw on physiological properties of the auditory system that are not unique to humans. Sensitivity to features other than voicing has been investigated, with varying degrees of success (see Jusczyk 1981). Others have tackled the difficult and important question of the relation between the infant’s ability to discriminate and language variation in use of particular features and their phonetic values.

Not all languages draw the boundary for perception of voicing at the same point that English does: some languages (for example, Thai) have a boundary at a point preceding the stop release. Thus the same stimulus will be perceived categorized as voiced or unvoiced by adult speakers, depending on the language they have learned (see figure 2.3). Such facts clearly illustrate the abstract nature of our mappings between phonetic feature values and the
speech stimulus. A sound that is categorized as voiced in English will be categorized as voiceless in Thai, if the VOT is sufficiently close to the stop release. Moreover, the switch-point that signals the voiced–voiceless contrast in English signals a different switch-point in Thai. Thai is a language that uses aspiration distinctively, producing three-way contrasts among stop sounds — for example, [h], [p] and [pʰ]. Late VOT coincides with aspiration, and is a sufficient cue for perception of sounds as aspirated by speakers of Thai. At roughly the same VOT that signals the switch from voiced to voiceless sounds for English speakers, the switch from voiceless to voiceless aspirate will be made by Thai speakers (see figure 2.4). Our phonological abilities thus include not merely the capacity to use different acoustic values as the locus for the switch from plus to minus values of the same feature, but also the capacity to use the same acoustic value as the switch-point for different features (a VOT of 30 milliseconds following the stop release signals the switch from plus to minus voice for English and from minus to plus aspiration for voiceless sounds in Thai).

It is generally agreed that there are typical switch-points for plus and minus feature values. One is at around the English voicing boundary of 30 milliseconds following the stop; another is somewhere in the region 20–50 milliseconds preceding the stop release (the precise value being a matter of dispute: see Lisker and Abrahamsen 1970; Eimas 1975). Other VOTs may be the locus of feature switches, but are arguably atypical. Spanish uses such an atypical value for the voiced–voiceless distinction, in the region between 20 milliseconds before and 20 milliseconds after the stop release.

Plainly, infants who display categorical perception cannot be credited with knowledge of the labels that adult speakers in different languages assign to different phonetic values. An appealing developmental hypothesis is that the
young infant is equipped with the perceptual apparatus to categorize according to a range of potential switch-points for feature values and learns the labels associated with particular points as he pieces together the phonological system of his language. When a particular potential switch-point for a feature is not used in a language, that point will no longer be a locus for categorical discrimination of sounds, and stimuli on either side of the boundary will come to be perceived as identical.

A separate, but related, developmental hypothesis concerns the claim that some potential switch-points for feature values are typical and used in many languages, while others are less frequently used. Developmentally, we might expect infants to be initially sensitive to typical boundaries and only later develop categorical perception for features whose values straddle less frequently used boundaries, such as the Spanish voicing boundary.

Thus there are two plausible developmental hypotheses: (1) children will show categorical perception for a range of distinctions, some of which will fade as the child acquires knowledge of language-particular facts; and (2) the earliest distinctions will involve typical switch-points, the perception of the less usual boundaries developing later. In general, early perceptual abilities will be refined and altered by experience with a particular language, and the appropriate phonological labels added into the perceptual categories the child retains and/or develops.

There is some support for this general picture. Lasky et al. (1975) found that four- to six-month-old infants in a Spanish-speaking environment discriminated between stimuli that had VOTs of 60 milliseconds and 20 milliseconds before the release of the stop closure and between stimuli with VOTs of 20 milliseconds and 60 milliseconds after the release. The same infants did not discriminate between stimuli at 20 milliseconds preceding the release and
20 milliseconds following the release. These infants picked up on the two differences that crossed boundaries used to demarcate speech-sound categories in many languages, but they were not sensitive to a difference of equal magnitude that less typically contains a boundary, and which in fact crossed the boundary used for voicing in the language they were to acquire (Spanish). Some studies of speech production have produced results that also support the idea that some boundary points are more basic than others; Gandour et al. (1986) found that three- and five-year-old children learning Thai more cleanly distinguished between voiceless aspirate and voiceless unaspirate stops, with a VOT boundary of approximately +30 milliseconds, than they did between voiceless unaspirate and voiced stops, for which there was considerable overlap in VOT values. This overlap was not found in Gandour et al.'s test for adults, who sharply distinguished between voiceless unaspirate and voiced stops, restricting voiceless unaspirates to a VOT range of approximately 0–+20 milliseconds. Another perception study that supports the view that the child’s early discrimination is relatively independent of experience is Streeter (1976), who found that infants in a Kikuyu language environment distinguished a boundary prior to the stop release for sounds for which voicing is not contrasted in Kikuyu.

But some results also argue that perception is very quickly tuned to the language-particular. Eilers et al. (1979) found that six- to eight-month-olds in a Spanish-speaking environment were sensitive to the Spanish boundary, whereas English-learning infants of the same age were not. Thus sensitivity to atypical boundaries may be triggered by language exposure before the child has any appreciable knowledge of the contrast, at the level of which features are distinctive in his language. The facts are not completely clear-cut at present, a number of experimental factors being confounded in the studies undertaken to date (see Jusczyk 1981 for discussion).

In sum, infant speech-perception studies constitute strong evidence for a particular kind of ability: the discrimination of speech segments according to phonetic values that are frequently employed in adult phonological systems. The development of language-particular perceptual boundaries and feature labels is not well understood.

2.4 Early Speech Sounds

Most children begin to produce recognizable words at some point in the second year (see chapter 4). Before that, children pass through a period in which speech-like sounds are produced, with no obvious link to words in the
adult language. Playful production of isolated consonant and vowel-type sounds (typical of four- to six-month-olds) is replaced at around six months by reduplicative babbling. The child produces series of consonant–vowel (CV) syllables, in which the individual syllables in each babbled series are identical or very similar to one another. At around ten months, this type of babbling gives way to syllable sequences with more varied members (different consonants and/or vowels) and a wider range of syllable types – VC and CVC in addition to CV (see, for example, Stark 1980). The next stage is the production of recognizable words, which may be preceded for some children by a 'silent period' in which babbling ceases (see Vihman et al. 1985 for a recent study).

The properties of babbled and first word speech have been the subject of many studies. Locke (1983, ch. 1) gives an extensive survey. When we look at the types of speech sounds children produce in the babbling and first word stages, and compare these to the inventory of sounds in the language around the child, the general picture is as follows. Babbling may include sounds not used in the language to which the child is exposed (for example, a child exposed to English may babble the velar fricative [X], mentioned above). None the less, there are clear preferences for certain sounds in babbling; stops, nasals and [h] are frequent in babbling, no matter what language the child is exposed to (see Locke 1983, p. 10); fricatives and liquids are generally avoided. Of the stops, voiced stops are more frequent in babbling than voiceless stops. A common change between early and later babbling is a decrease in the use of back (velar) stops. First words tend to be composed of a narrow range of sounds, typically front voiceless stops, nasals and the vowel [a]. (That is, the child's pronunciations tend to favour these sounds, although this may involve a distortion of the corresponding adult word.) As the child progresses with real-word speech, more sounds are added to his inventory, so that the range of sounds he produces more nearly matches that of the language he is learning. Use of speech sounds in babbling and early speech thus broadly fits an hour-glass pattern; sounds used in babbling (for example, velar stops) may drop out in early speech and then be reintroduced.

Jakobson (1968) focused on babbling and early speech in the context of the distribution of sounds in languages of the world. He proposed that there are regular relationships between the distribution of speech sounds cross-linguistically and the order in which different sounds are acquired. Front voiceless stops, nasals and the vowel [a] are found in virtually all languages. Drawing on many observational studies of child language, Jakobson proposed that such extremely frequent sounds were the first to be acquired. Other sounds occur less frequently in languages of the world and the presence of these less frequent sounds can be used to predict the presence of more frequent sounds. So, for example, back (velar) stops are less frequent than front stops
and the occurrence of back stops is a good predictor of the presence of front stops in the world's languages: a language may have front stops without back stops, but not vice versa. Similarly, fricatives as a class occur less frequently than stops (a language may have stops without having fricatives, but the reverse situation does not occur). These frequency facts mirror acquisition facts: front stops are mastered before back stops and stops are mastered before fricatives. Some sounds are rare in the sound inventory of languages; an example is the English vowel [æ], which is a sound that occurs with high frequency in English, but is rather rare in the world's languages. Jakobson observed that cross-linguistically rare sounds are generally acquired late.

It has become usual to refer to frequency facts for the distribution of sounds in terms of markedness. Markedness is a term used to refer to the extent to which a phenomenon is 'normal' in language systems. The most frequent sounds are referred to as maximally unmarked; sounds whose presence can be used to predict the presence of other sounds are marked, relative to the sounds of which they are predictors. Rare sounds can be referred to as the most marked sounds.

A general goal of linguistic theory is to try to find out whether frequency (markedness) observations can be accounted for in terms of principles of linguistic structure. Can Jakobson's markedness observations be recast in terms of a principled organization of sound systems? Is it the case that phonetic features are intrinsically related in such a way that some sounds logically have priority in languages of the world and also must be acquired before others in language development? At present, the evidence does not exist to support such a strong position. It is certainly possible to formalize frequency observations with respect to sound systems. If segments are characterized in terms of plus and minus specifications on features, markedness values for segments can be viewed as default, or normal, specifications for the values of a feature in combination with other feature values. So the default, unmarked specification for a vowel (a [+consonantal] sound) would be [+low], the specification for the most common vowel, [a]. Other specifications will have a 'cost', which may be expressed as an algorithm that translates particular combinations of features into numerical values (see Kean 1976/1980 for a system of that type). But there is nothing in the algorithm itself that mandates that the maximally unmarked vowel sound is [+back], and so forth. Markedness values established in such a way are essentially stipulative, encoding the distributional facts that happen to hold in human languages, child and adult.

Another question with respect to babbling and early speech is the influence of factors other than markedness values on the patterns observed. Two possible sources of influence other than markedness values are articulatory control and the speech sounds to which the child is actually exposed. Articulatory factors may well play a role. The predominance of voiced stops over voiceless stops in babbling is a counter-example to the claim that unmarked sounds
predominate; voiceless stops are more frequent in languages of the world than their voiced counterparts, and emerge earlier than voiced stops in early, non-babbled, speech. The predominance of voiced stops in babbling may be a consequence of insufficient articulatory control for the production of voiceless sounds, to produce which the vocal cords must be held apart.

The influence of speech sounds the child hears is not clear, beyond the obvious fact that the child ultimately acquires those sounds to which he is exposed. An early hypothesis was that there was a shift during babbling towards the sounds of the language to which the child was exposed (Brown 1958). There now seems little to support this idea. Adults do not generally find it easier to identify the language background of older babblers than younger babblers (Locke 1983, pp. 13ff; but see also Vihman et al. 1985). Possibly the illusion of a shift towards sounds of the language the child is learning may be created by a shift towards use of unmarked sounds (as, for example, when velar stops come to be used less frequently in latter babbling), which are those most likely to be found in any language.

2.5 Feature Acquisition

Even if markedness values that express the preferences for some sounds in babbling and early speech have not to date been made to follow from intrinsic properties of the phonetic feature system, that does not mean that features are not useful and necessary in describing children’s early production of sounds. A child may impose patterns on her babbling or early speech that are neatly characterized in terms of features. Gruber (1973) showed that in later babbling one child in an English-speaking environment preferred sequences in which syllables with alveolar consonants preceded sequences with labials or velars. Sequences such as that in (5)

(5) [do te ma ga]
   alveolar labial/velar

were preferred, and sequences such as that in (6)

(6) "[do ma te ga]"

were avoided. Alveolar sounds are produced with an obstruction by the tip or blade of the tongue at the alveolar ridge (the hard ridge directly behind the
teeth); the feature [± coronal] distinguishes alveolars from labial and velar sounds, where the obstruction is more peripheral. Alveolars are [+ coronal] and labials and velars are [− coronal]. The sequencing constraint imposed by the child Gruber studied can be described as a constraint that all [+ coronal] sounds precede [− coronal] sounds. (Gruber expresses the generalization in terms of the feature [± grave], a feature motivated by the fact that labials and velars share some acoustic properties that distinguish them from alveolars; Jakobson et al. 1952.)

Sometimes a child broadens her repertoire of sounds in a way that fits exactly with the introduction of a new phonetic feature. For example, when a child begins to make use of the voiced/voiceless distinction in stops she may do so for all the stops she uses. (Note that although Eimas's study demonstrated that even infants can distinguish voiced from voiceless stops, it is some time before children can use this distinction in producing and understanding the language around them.) When a child begins to use voicing for stops, a three-way distinction [p, t, k] may become a six-way split [p−b, t−d, k−g], as we would expect if the feature [± voice] is suddenly introduced into the set of features she controls. Other children, however, introduce contrasts piece-meal – for example, voicing labial or alveolar but not velar stops. This type of case is not problematic for the notion that the child is learning to make use of feature contrasts (contrary to what is sometimes assumed; see, for example, discussion in Dale 1976). Adult languages also on occasion do not impose the full range of possible contrasts allowed by the features needed to characterize their inventory of sounds. For instance, in English we need the feature [± continuant] to distinguish stops from fricatives, but there is no velar fricative in English, so this feature is not fully utilized. Piecemeal addition of contrasts by children represents acquisition stages analogous to these adult systems. For adult systems, feature representations are amply motivated. Not only are features an important descriptive device in characterizing speech sounds, they also have a critical role in characterizing the systematic patterning of sounds both within and across languages – phonological systems. The next section deals with children's developing phonological system.

2.6 Child Phonologies

Phonological systems have traditionally been viewed as comprising rules for two types of phenomenon: segmental and suprasegmental. Segmental rules
affect the presence of and precise phonetic feature values associated with individual consonant and vowel segments. Learning the segmental system of a language will involve working out which properties of the sound segments that make up words in the language are non-predictable and distinctive (used to distinguish words from one another) and which properties can be predicted by rule. Non-predictable properties will be entered in the lexical entry for the word; predictable properties will be spelled out by phonological rules that work on and change the basic lexical entry for the word. To take the example of aspiration given above, a child learning English must work out that the aspiration that he hears on the [k] in a word like [kæt] is predictable from the word-initial position of the [k], and so need not be part of the lexical entry for that word. Aspiration can be spelt out by phonological rule. A child learning Thai will have to come to the opposite conclusion. In Thai aspiration is not predictable, does distinguish between the meanings of words, and must be entered as a feature in the lexical entries for Thai words.

Suprasegmental rules concern properties of pronunciation that typically affect constituents larger than the segment, particularly stress, intonation and tone. More details of these phenomena are given below.

We will assume a model of grammar of the following form. Rules of phonology interact on a structural, syntactic representation of the sentence. The lexicon specifies unpredictable aspects of pronunciation (the fact that the

![Diagram](image)

*Figure 2.5*
word 'cat' comprises three segments, the first of which is a voiceless velar stop, etc.), and rules of syntax will provide a specification of the syntactic environment in which a word occurs, which may affect pronunciation in several ways. In terms of the model of grammar we will use in later chapters, rules of phonology apply to the S-structure of the sentence: this is approximately a representation in which the linear string of words in a sentence is organized in an abstract, hierarchical structure that represents some, but not all, of the logical and referential relationships between words and phrases in the sentence (see chapter 4). The output of phonological rules will be a representation called phonetic form (see figure 2.5).

2.6.1 Segmental Rules

Some rules of segmental phonology appear to be very quickly incorporated into a child’s pronunciation: for example, aspiration of initial voiceless stops in English seems to be rapidly mastered. But children’s pronunciations are far from error-free. Children may make frequent and systematic errors vis-à-vis the adult target words.

Many recurrent error types in child speech show a pattern corresponding to phonological rules in languages of the world. For example, children not uncommonly de-voice segments in final position, pronouncing ‘bag’ as ‘bak’ [bæk]; such a devoicing rule is found in many languages (for example, German), although it is not a rule of English segmental phonology. Or they may move the place of articulation forward for stops and other segments, pronouncing 'key' ([kɪ]) as ‘tea’ ([tɪ]), and so forth; again, processes of this kind are found in adult phonological rule systems. Children also frequently simplify consonant clusters, deleting consonants or introducing a vowel to break up the cluster. Table 2.3 gives a list of frequent types of child mispronunciations, with examples.

What is the status of children’s mispronunciations? There are at least three separate possibilities. Child mispronunciations could reflect (1) incorrect lexical representations on the child’s part, with a non-adult lexical entry for the word in question; (2) non-adult phonological rules operating to distort adult-like lexical representations; or (3) non-systematic errors of articulation.

It is not always clear which of these three possibilities is the correct way to look at a given child error, but it is evident that in many instances the child does have in his head something like the correct adult form, even when he does not pronounce the word correctly (consistent with (2) and (3)).
Table 2.3  Some typical child pronunciation errors in the second and third years

<table>
<thead>
<tr>
<th>Substitution processes</th>
<th>Child pronunciation</th>
<th>Age (yrs; mths)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stopping (a fricative is replaced by a stop)</td>
<td>see ti:</td>
<td>2; 9</td>
<td>Smith (1973)</td>
</tr>
<tr>
<td>Fronting (the place of articulation is fronted, with velar and palatal consonants being replaced by alveolars)</td>
<td>shoe zu</td>
<td>2; 0</td>
<td>Velten (1943)</td>
</tr>
<tr>
<td>Gliding ([w] or [j] is substituted for a liquid)</td>
<td>leg jek</td>
<td>2; 1</td>
<td>l</td>
</tr>
<tr>
<td></td>
<td>ready wedi</td>
<td>2; 1</td>
<td>l</td>
</tr>
</tbody>
</table>

Assimilation processes (a sound becomes more similar to an adjacent sound)

<table>
<thead>
<tr>
<th>Voicing (consonants tend to be voiced preceding a vowel and devoiced at the end of a syllable)</th>
<th>Child pronunciation</th>
<th>Age (yrs; mths)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>paper be:bo</td>
<td>2; 3</td>
<td>Smith (1973)</td>
<td></td>
</tr>
<tr>
<td>pig bik</td>
<td>1; 5</td>
<td>l</td>
<td></td>
</tr>
</tbody>
</table>

Consonant harmony (Consonants tend to assimilate in words with the structure C1VC2(X))

<table>
<thead>
<tr>
<th>Diphthong assimilation</th>
<th>Child pronunciation</th>
<th>Age (yrs; mths)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>duck gˌʌk</td>
<td>1; 7</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>tickle gˌɪgju</td>
<td>2; 2</td>
<td>Smith (1973)</td>
<td></td>
</tr>
<tr>
<td>tub bˌæb</td>
<td>(no age given)</td>
<td>Menn (1975)</td>
<td></td>
</tr>
</tbody>
</table>

Progressive vowel assimilation (an unstressed vowel will assimilate to a preceding vowel)

<table>
<thead>
<tr>
<th>Diphthong assimilation</th>
<th>Child pronunciation</th>
<th>Age (yrs; mths)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>bacon bʊ:du</td>
<td>2; 0</td>
<td>Velten (1943)</td>
<td></td>
</tr>
<tr>
<td>flower fɑ:wa</td>
<td>2; 0</td>
<td>Velten (1943)</td>
<td></td>
</tr>
</tbody>
</table>
Moreover, there is evidence that some mispronunciations are the result of non-adult phonological rules (consistent with (2)), rules that sometimes attest to considerable powers of generalization and abstraction at the phonological level.

One clear type of evidence of adult-like lexical forms is the fact that children can recognize their own mispronunciations as deviant (adult: ‘Did you say “wellow”?; child: ‘No, I said “wellow”’). Another kind of argument for adult-like lexical entries is the fact that a systematic account of child articulations can be given if the adult forms are also the child’s representations, at an underlying level. Some examples of plausible child rules operating on adult-like underlying forms will illustrate this point and also the child’s rule-forming capacity.
In a detailed study of the phonological development of one child, N. Smith (1973) takes the position that the adult form is always the form in the child’s mental dictionary, and mispronunciations are a matter of realization rules (Smith’s term) that convert the adult form to the child’s pronunciation. (These output rules can be regarded as non-adult phonological rules that must be eliminated as the child moves towards adult phonology.)

Smith proposed that between the ages of approximately two years two months and two years four months the child was using the rules in (7), among others (the rule numbers are taken from Smith):

(7) Rule 1: If a nasal is followed by a voiceless consonant, the nasal is deleted.
Rule 2: If a nasal is followed by a voiced consonant, the voiced consonant is deleted.
Rule 3: A coronal stop is velarized before /s/.
Rule 4: An unstressed vowel is raised (made higher) and backed before /l/.
Rule 6: /l/ is deleted at the end of a word.
Rule 13: /h/ is deleted.
Rule 21: An alveolar consonant is deleted after another consonant.
Rule 25: All voiceless segments are voiced.

These rules can be used to account for the child’s pronunciation of words such as ‘empty’ and ‘handle’. The path from the adult form to the child pronunciation as given by Smith is as shown in (8), where the adult pronunciation is taken as the starting-point and the rules are applied as shown to change the adult form into the child pronunciation:

(8a)

<table>
<thead>
<tr>
<th>Adult form:</th>
<th>‘handle’</th>
<th>/hændəl/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child rules:</td>
<td>Rule 2</td>
<td>/h ænəl/</td>
</tr>
<tr>
<td></td>
<td>Rule 3</td>
<td>/h ænəl/</td>
</tr>
<tr>
<td></td>
<td>Rule 4</td>
<td>/h ænəl/</td>
</tr>
<tr>
<td></td>
<td>Rule 6</td>
<td>/h æŋu/</td>
</tr>
<tr>
<td></td>
<td>Rule 13</td>
<td>/œŋu/</td>
</tr>
</tbody>
</table>

[œŋu] = child pronunciation
Phonological Acquisition

(8b)
Adult form: 'empty' /empti:/
Child rules:

Rule 1 /epti:/
Rule 21 /epi:
Rule 25 [ebi:] = child pronunciation

The child's rules Smith proposes are quite general, applying to all or almost all the relevant forms in the child's vocabulary. Two related points can be made concerning Smith's analysis. First, for the analysis to work the child must be credited with some knowledge of the correct adult forms since the child's pronunciations are dependent on features of the adult form that are not present in the child's pronunciation. For example, Rule 3 states that a coronal stop becomes velar before /l/; that is, in adult words without a following /l/, /n/ will not be pronounced as /n/. Yet the /l/ is not present in the child's pronunciation of words such as 'handle', since Rule 6 deletes /l/ at the end of a word. The fact that Rule 3 includes a condition that is based on the adult form entails that the child has some knowledge of that form, which is not directly revealed in the child's pronunciations. Similarly, Rule 6 requires the /l/ as the condition for raising and backing the second vowel in 'handle'.

The second point is that the child's rules must apply in a particular order. For example, the child's Rule 3 (and Rule 4) must apply to the adult form before Rule 6, since Rule 6 deletes the element /l/ that is the condition for application of Rule 3 (and 4). The ordered application of phonological rules is a characteristic of rule application in theories of adult phonology; thus it appears that although the child adopts rules that are not part of the adult phonological system of his language, he applies those rules in a way consistent with the constraints on rule application in adult phonologies.

Smith's analysis therefore entails that the child has available something identical or similar to the adult form and applies his own rules to that form in an ordered way to arrive at his own pronunciation. These basic points are widely accepted in the literature on phonological acquisition, although the very strong position that Smith takes – that the child has in his head a representation that is identical to the adult's pronunciation – has been challenged. Braine (1976) pointed out that there is a plausible perceptual basis for some of the phenomena Smith captures with his realization rules. For example, Rules 1 (which deletes a nasal before a voiceless consonant) and 2 (which deletes a voiced consonant after a nasal) may alternatively be accounted for in terms of perceptual error. Vowels in English are lengthened before a voiced consonant. Acoustically, vowels carry cues for the perception
of a following nasal, and the lengthening of the vowel as a consequence of a following voiced consonant may thus cause a nasal in a V-nasal-C sequence to be perceived, where it would not be perceived if the consonant were voiceless and the vowel not lengthened (see also Macken 1980).

In addition to questions concerning the role of a perceptual filter, there is continuing debate over whether the child has only one lexical representation for a word that undergoes a set of phonological rules or two lexical representations, one accessed in comprehension and one that forms the basis for production of words; this point is returned to briefly below (section 2.7).

It is important to note that children's non-adult pronunciations are not a response to absolute inability to produce certain sounds. It is frequently the case that where a sound is eliminated by a child rule, the same sound will be introduced by another rule in the child's phonology. A well-known example from Smith's (1973) study concerns the pronunciation of alveolars and velars. For a time, the child velarized alveolar stops before [l], pronouncing the word 'puddle' as 'puggle' ([pəgəl]). At the same period, he changed non-final [z] to [d], pronouncing 'puzzle' as 'puddle' ([pʌdəl]). The mispronunciation of adult 'puddle' is then clearly not rooted in some motor deficit, since the child produces the adult form in his version of adult 'puzzle'. Stampe (1972, ch. 1) lists many examples that make the same point.

### 2.6.2 Suprasegmentals

Recent phonological theory has argued that phonological representations are multi-dimensional. Stress, intonation and (for languages that use pitch height distinctively) tone are all 'suprasegmental' phenomena, and all are characterized in recent phonological theory by representations that are distinct from the phonetic feature representations we have discussed so far. The specifications for these suprasegmental phenomena are connected to the string of consonantal and vowel segments by rule, creating a multi-dimensional representation.³

**Stress** Stress rules must account for the fact that a particular syllable or syllables in a word or phrase is/are perceived as more prominent than the rest. The acoustic basis for prominence is a complex matter, with several factors (intensity, length, pitch) contributing to what we perceive as prominence (Lehiste 1970). The discussion below is concerned with the rules for stressing words only, not phrases. The type of representation that is now widely used for the stress pattern of words is a representation in terms of a hierarchical tree structure in which the branches are labelled as strong or weak. The syllable of a word that has maximum stress will be the syllable that is hooked to a strong branch, or chain of strong (s) branches. For example, the stress
on the words 'usurp' and 'develop' can be represented in terms of stress trees of the (simplified) form in (9):

\[
\begin{align*}
(9) \quad \text{usurp} & \quad \text{develop} \\
\text{w} \quad \text{s} & \quad \text{w} \quad \text{s} \quad \text{w} \\
\text{s} & \quad \text{s}
\end{align*}
\]

Tracing a path from the root of the tree up through a chain of one or more s-nodes will lead to the syllable with primary stress, indicated by the stress mark ' over the vowel. (The representations here are adapted from Hayes 1982, whose general approach to stress rules builds on Liberman and Prince 1977.)

Stress rules must construct the correct stress trees for a given language. Stress rules for words offer a good example of parameters in the domain of phonology. The rules for stressing words in a language tend to be 'anchored' either to the right or the left of a word. There is thus an anchoring parameter for which languages will choose a value. Some languages have rather simple systems – for example, placing stress on the first or second vowel from the left or right of the word’s edge (see Hammond 1988 for a recent survey).

Other languages, including English, have a relatively complex system that pays attention not merely to distance from the edge of the word, but also to the nature of the material at the word’s edge and the syntactic category (noun, verb, etc.) of the word that is being stressed. The child’s task in learning the stress rules for words in his language will be to figure out the direction in which stress is determined and the exact nature of the rules. Take the words ‘usurp’ and ‘develop’ given above. The stress pattern on those two words taken alone is compatible with either of the following two hypotheses:

*Hypothesis 1*: English is a left-anchored language, with a general rule: stress the second syllable from the left edge of the word.

*Hypothesis 2*: English is a right-anchored language, in which stress is placed sometimes on the syllable immediately at the right edge and sometimes on the second syllable from the right edge.

The first hypothesis is clearly simpler, but it is incorrect. Hypothesis 2 is the correct hypothesis for English, and the child will have to work this out, and also to learn what the conditions are that determine when stress is placed on the rightmost syllable of the word ('usurp') and when stress is placed further back in the word ('develop'). Two factors affecting whether or not a word is end-stressed in English or not are: (1) the 'weight' of the final syllable and (2)
the syntactic category of the word. The weight of a syllable depends on the number of consonants that follow the vowel and the quantity of the vowel (roughly, whether the vowel is long or short). The exact characterization of vowel length is a relatively complex matter. It is sufficient here to note that length distinguishes between pairs of vowels in English, such as the long vowel [i] (the vowel in ‘beet’) and the short vowel [ɪ] (the vowel in ‘bit’). In English, a syllable is heavy if it has a long vowel and/or has a final consonant.

Although the exact rules are complex, the broad effect of the weight of the word-final syllable in English is that a heavy syllable in final position is likely to prevent stress from moving towards the middle of the word. For adjectives and verbs, the weight of the final syllable is calculated ignoring the last consonant. Stress is placed on the last syllable of ‘usurp’ because the syllable has two consonants at the end, and is heavy even if the last consonant is ignored; similarly, stress goes on the last syllable of the adjective ‘discreet’ because the vowel in the last syllable is long, and a long vowel always means a heavy syllable. In a word such as ‘develop’, the vowel in the last syllable is short and followed by only one consonant, and stress can skip over the last syllable.

Syntactic category adds complication to the system. For nouns, there is a tendency for stress to be placed one syllable to the left of where it would be placed for verbs and adjectives. Thus we have nouns such as ‘insect’, with main stress on the first syllable despite the fact that the second syllable ends in two consonants. Although there is a good deal of complexity in the system, and it is recognized that the generalizations found are a matter of tendencies rather than absolute rules, the generalization that stress goes to the left in nouns as opposed to verbs is a real one, and shows up clearly in contrasts such as that in (10):

(10) convict convex
     (verb) (noun)

How do children fare in learning the stress system of their language? It is notable that even in a language such as English, which has a complex stress system, children do not seem to make many errors with stress in their spontaneous speech. (Those errors that are made may be related to syllable structure; see Smith 1973 and Klein 1984.) Several studies support the idea that stress systems are quite easily mastered.

In computer simulation of stress acquisition, Dresher and Kaye (1986) have shown that various stress systems (left-anchored or right-anchored) can be ‘learned’ quite efficiently by a machine exposed to basic data (words in the language to be learned). The machine is equipped with prior knowledge of the factors that affect stress systems in the world’s languages, such as anchoring direction and syllable weight, and formulates hypotheses based on this
knowledge and the input data. This is the type of knowledge with which the child would be expected to tackle the language learning process, since these ingredients of stress systems are presumably part of universal grammar.

In experimental work, P. Smith et al. 1982 (summarizing results of Smith and Baker 1976 and Groat 1979) have argued that at least by age seven, children have grasped the elements of the English stress system — in particular, the effects of syllable weight and word class described above. Both child and adult subjects were asked to read aloud nonsense words, which were presented in contexts such as those in (11), which made the nonsense word unambiguously a noun or verb:

(11) The nuvit was made in a factory (noun context)
The man had to nuvit the tractor (verb context)

The child subjects were seven-year-old schoolchildren in Edinburgh, Scotland; the adult subjects were British university students. All of the test words had two syllables; they had either a single consonant in final position (as in the ‘nuvit’ example) above, or two consonants in final position (for example, ‘rafust’). Length of vowels is not represented in a consistent way in English spelling, so whether the final vowel was interpreted as long or short was a matter of how the subject chose to pronounce the word. A word such as ‘nuvit’ could be pronounced with either a long or short final vowel. Table 2.4 shows the percentage first-syllable stress according to number of final consonants (one or two), quality of the final vowel (long or short) and category of the word (noun or verb). The table shows that children have essentially the

<table>
<thead>
<tr>
<th>Number of final consonants</th>
<th>Quality of final vowel</th>
<th>Example of nonsense word</th>
<th>Example of similar English word</th>
<th>Children Noun</th>
<th>Children Verb</th>
<th>Adults Noun</th>
<th>Adults Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Short</td>
<td>nuvit (nuvit)</td>
<td>Edit</td>
<td>89</td>
<td>78</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>Long</td>
<td>nuvit (nuvit)</td>
<td>Discreet</td>
<td>65</td>
<td>38</td>
<td>71</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Short</td>
<td>rafust (rafust)</td>
<td>Distrust</td>
<td>87</td>
<td>56</td>
<td>50</td>
<td>25</td>
</tr>
</tbody>
</table>

* Smith et al. use the terms tense and lax for long and short.

* The transcriptions are as given by Smith et al. The vowel transcribed as [i] would appear to be the vowel we have designated with [i].

Adapted and abbreviated from Smith et al. (1982), table 4.
same pattern of responses as adults. Both are more inclined to put stress on
the first syllable if the word is a noun than if it is a verb; if it has a short final
vowel rather than a long final vowel; and if there is one consonant at the end
of the word rather than two (for children, the last holds only in the case of
verbs). Thus the seven-year-olds in this study had extracted from the real
words they had heard the essential elements of the rules of the stress system
of their language. A recent study by Hochberg (1988) uses both experimental
data and spontaneous speech to argue a similar point with respect to Spanish,
using subjects as young as three years.

Intonation

Intonation and tone are two further phenomena for which
phonological theory posits levels of representation separate from the string of
consonant and vowel segments. Specifications of pitch height critical to the
intonation patterns of a language such as English and for tone values in a
language such as Chinese (see below) will be spelled out in terms of a sequence
of pitch segments, linked to the string of vowel and consonant segments that
make up a word or phrase by rule. In English, the normal intonation for a
statement is a rise followed by a fall in pitch:

(12) \[ \ downsweep \ downsweep \ downsweep \ downsweep \]

John broke the balloon

This can be translated into a representation such as (13) where the pitch
melody is (minimally) schematized as a sequence of \textit{H L} (high; low) pitch
segments, which can be linked to the consonant and vowel string by rule
(Selkirk 1984, ch. 5 and references therein):

(13) \begin{align*}
\text{H} & \quad \text{L} \\
\end{align*}

John broke the balloon

The exact positioning of the pitch peak (H) may vary, but the general H-L
contour will ordinarily be present in declarative sentences. Questions in
English (and many other languages) have a different, L-H pattern, as shown in
(14)--(15):

(14) \[ \ downsweep \ downsweep \ downsweep \ downsweep \]
Was the skunk black?

(15) \begin{align*}
\text{L} & \quad \text{H} \\
\end{align*}
Was the skunk black?
On this approach to intonation, the pitch melody is thus represented as a sequence of segments at a level separate from the sequence of consonant and vowel segments.

Studies of early child language show that children know the value of these basic intonation patterns early on. In early stages of development, children learning English frequently omit inversion of subject and verb ('The skunk was black?' – 'Was the skunk black?'), signalling a question with rising intonation only (see Klima and Bellugi 1973). Moreover, the speech of some children offers some support for a general approach to intonation in which the pitch melody is represented as an autonomous level, distinct from the sound-segment level. Peters (1977) studied the speech of one young child who appeared to have two distinct speech modes. At around one and a half years, this child produced on the one hand fairly intelligible one-word utterances and on the other hand longer utterances with low intelligibility. These latter utterances, Peters argues, preserve the intonation contour of the target utterance, omitting or severely distorting the segmental content. For example, the (initially unintelligible) utterance

(16) [á lor ri gò mu nyai]

was pronounced with an adult-like intonation tune for the intended sentence, 'I like to read Good Moon Night' (Good Night Moon).

This kind of behaviour makes sense within the picture of intonation sketched above, where the intonation tune is separate from the segmental string. At one and a half or two years the child may have a quite solid command of vocabulary, syntax and prosody, but calling up those different features of his knowledge in a single utterance may be simply beyond his capacities. So he may plan a fully-fledged utterance, complete with syntax, words and appropriate intonation, but be able to execute only one part of that plan – either one or two individual words are carefully articulated or the overall intonation and a few features of the words are realized at the expense of intelligibility. The partial independence of intonation and the segmental (CV) representation of words can be explained in terms of the partitioning of the child's energy in speech production. (Peters 1977 presents a somewhat different account.)

Tone In tone languages, pitch is used to distinguish words and morphemes and is thus a lexical property. As an example, Mandarin Chinese uses four distinctive tones: high-level, rising, falling and dipping. These terms reflect the pitch levels associated with each tone; high-level tone is a relatively high pitch, rising tone is a shift from a lower to a higher pitch level on a single vowel,
falling tone is a shift downwards in tone and dipping tone is a slight fall in tone followed by a rise. The forms in (17) illustrate the way in which tone can distinguish between otherwise identical words:

(17)  HIGH     bā     ‘eight’
RISING   bá     ‘to pull’
FALLING  bà     ‘a harrow’
DIPPING  bā     grammatical marker for object

(The diacritics for the different tones are from the Pinyin romanization system.) In addition, Mandarin has a ‘neutral’ tone, which occurs in unstressed syllables and has different level tone values (mid, half-low, low), depending on the tone of the preceding syllable. Tone languages differ with respect to the level tones (high, mid, low) and contour tones (rising, falling, etc.) they use, and the extent to which lexically assigned tones (tone values that are part of the lexical entry of the word) are adjusted by phonological rule.

A standard approach to tone systems in recent linguistic analysis is to treat tone as a specification of pitch height in terms of pitch segments, just as in the treatment of intonation in the preceding section; contour tones can be represented as the assignment of a string of two or more pitch height segments to a single vowel. The Mandarin words in (17) can be represented as in (18):

(18)  bā  bá  bà  bā
      H   L   H   H   L   M   L   H

‘eight’  ‘to pull’  ‘a harrow’  object marker

(H = high pitch; L = low pitch; M = mid pitch)

Both general constraints and language-particular rules will govern the linking relations between tone segments and vowels (see Goldsmith 1976 for an early treatment).

Li and Thompson (1978) and Clumeck (1980) review studies on the acquisition of tone. Overall, studies of children acquiring a number of different tone languages reveal little evidence of any special difficulty with tone-to-word mappings, which generally seem to be mastered before the segmental system is fully acquired. Demuth (1989a) uses data from children acquiring Sesotho to argue that from an early stage (two years) children distinguish tone from stress.

One particularly interesting finding concerns the acquisition of contour
tones. Although there is some variation concerning which tones are mastered earlier than others, at least some children disfavour contour tones. For example, Li and Thompson (1977) report that one child learning Mandarin Chinese replaced contour tones with level tones of different heights, using a low tone for falling and dipping tones in adult forms and a mid or high tone for rising tones in adult forms (in the examples Li and Thompson cite). Within the type of tone system sketched above, this type of behaviour is easily characterized. One has simply to say that the child imposed a ban on many-to-one mappings between tone segments and vowels; he does so by systematically adjusting the adult representations above, stripping away the left of the adult tone segments in falling and rising tones and both peripheral segments in the dipping tone.

A system of representation with tone segments thus permits the child’s behaviour to be perspicuously described. Difficulty with contour tones can be accounted for in terms of a plausible restriction – one-to-one mapping between levels. Moreover, systematic simplification of contour tones is another type of evidence that the child may have an adult-like representation for a form, even though his pronunciation deviates from the adult’s pronunciation. We cannot explain why a child produces a low tone for a falling or dipping tone and a mid or high tone for a rising tone unless we credit the child with having perceived a difference between these contour tones. The child’s treatment of the different tones can be described in terms of a systematic adjustment to the adult’s representation of the word.

Syllable structure and sonority Syllables have a hierarchical internal structure whose organization is sensitive to the feature composition of the component parts. Much linguistic and psycholinguistic evidence points to the syllable in English having a structure in which there are two primary units: an onset and a rhyme. The former comprises the initial consonant (or consonant cluster) and the latter comprises the vowel or other syllabic element, together with any syllable-final consonants. Thus a monosyllabic word such as ‘drunk’ has approximately the internal structure shown in (19):

\[ (19) \]

\[
\text{syllable} \quad \overline{\text{onset}} \quad \text{rhyme} \\
\text{nucleus} \quad \text{coda} \\
d \quad r \quad \lambda \quad n \quad k
\]

(See Cairns and Feinstein 1982 for one example of a more detailed account of the internal structure of syllables.)
Although there are language-specific variants, the ordering of elements in both the syllable as a whole and its internal units generally conforms to the dictates of a 'sonority hierarchy'. Sonority is the acoustic property we perceive as resonance. The less obstruction there is in the vocal tract, the more vowel-like or sonorant a consonant is. Stops are the least sonorant sounds, fricatives somewhat more sonorant, nasals and liquids more sonorant still. The normal case with respect to syllable organization is that segments on the periphery of the syllable are the least sonorant, with progressively more sonorant elements towards the nucleus of the syllable. The dictates of this hierarchy are followed in the example 'drunk' above, where the stop [d] precedes the liquid [r] in the onset, a vowel occupies the nucleus and the nasal [n] precedes the stop [k] in the coda. There are, however, exceptions to this general organizational pattern. For example, English permits words such as 'skunk', where the initial [sk] cluster contains a fricative followed by a stop, i.e. a sequence in which sonority decreases on the path to the nucleus.

Children's performance shows a sensitivity to different syllable types that reflects the relative frequency of the various syllable types in languages of the world and their internal organization. CV is the most common type of syllable in languages of the world (a fact that has been unified with the dictates of the sonority restrictions on syllable organization by Clements 1988; see Martohardjono 1989 for discussion with respect to language acquisition). Children initially eschew clusters of consonants; they babble primarily CV and then CV and CVC syllables, and in their first words frequently delete elements from consonant clusters or break up clusters with vowels to make syllables that conform to a CV pattern (see table 2.3 for examples). Experimental tests carried out with older children confirm sensitivity to hierarchical structures such as that above: Treiman (1985) found that children aged eight find it easier to perform a task that requires them to replace two sound segments in nonsense words when the replacement respects the onset–rhyme boundary of the stimulus item than when the onset–rhyme boundary is crossed by the replacement.

When the child begins to produce consonant clusters, there is evidence of sensitivity to the sonority hierarchy. For example, the child studied by Smith (1973) at a particular stage reduced initial triconsonantal clusters of the form [str] to two consonants. Of the three logically possible reductions – [st], [tr] and [sr] – he used only the last two. Thus a word such as 'straw' was pronounced 'traw' or 'strow', but not 'staw'. The reduced forms uttered by the child were those that conformed to the sonority hierarchy restriction on syllable organization – in the syllables 'traw' and 'strow' the ordering least-to-more sonorant is observed; the syllable 'staw' (not used by the child) violates that ordering. The effects of a general property of syllable structure rather than language-particular forms are clear in this example; in his cluster
simplifications, the child used a cluster that does not occur in adult English (English has no words beginning with [sr]) and avoided a sequence ([st]) that is found in English (in words such as 'sting'), beginning with the offending [st] cluster.

2.7 Problems and Ideas

The thrust of all the above sections is that the child's phonological abilities reflect the units, levels of representation and (as far as is known) the constraints that characterize adult phonological systems. However, it is not always the case that the child's behaviour mirrors the general patterns of adult phonologies. One example that has been frequently discussed is consonant harmony — agreement of consonants for features such as nasality and manner of articulation. Such harmony processes are quite rare in adult languages, but quite common in child pronunciations (see table 2.3 for some examples). Why should this be?

An idea that has recurred in various forms is that consonant harmony reflects the child's utilization of levels of phonological structure (such as the syllable) that are not normally the 'home' of the phonetic features that characterize the harmony. (See Menn 1978, Iverson and Wheeler 1987 for pertinent discussion). Thus, for example, a child who pronounces 'dog' as [ɡag] and 'coat' as [kok] may be hypothesized to be associating the feature [− anterior] with the entire syllable (equivalent in these examples to the word), as shown in (20):

(20) \[
\begin{array}{c}
\text{syllable} \\
[− \text{anterior}] \\
\text{onset} \quad \text{rhyme} \\
| \\
\text{g} \quad \text{a} \quad \text{g}
\end{array}
\]

All of the segments to which the feature [− anterior] is potentially relevant will be so characterized. Iverson and Wheeler propose an analysis along these lines. This type of analysis is appealing to the extent that (1) it provides a means of characterizing a phenomenon in child pronunciation that is odd, vis-à-vis the characteristics of adult languages and (2) the explanation provided ties in with levels of representation necessary in adult phonologies. However, the idea is unsatisfactory to the extent that in a sense it does little more than re-state the problem by notating the syllable node with particular features.