17 The Acquisition of Phonology

KATHERINE DEMUTH

1 Introduction

The field of phonological acquisition has changed and developed over the years, in tandem with various developments in phonological theory. Like the field of phonology itself, the study of phonological acquisition began with an examination of segments and words (Yeni-Komshian *et al.* 1980). Much of this work found a large amount of individual variation in children's early productions. This led to a more cognitive approach to phonological acquisition, where issues of individual strategies predominated, with little in the way of predictions about the course of phonological development. To be sure, the field was also small, and there was a limited amount of data available for investigating issues of phonological development. Over the past 15 years this situation has begun to change, and there has emerged a growing amount of literature examining the acquisition of phonological structure at the higher levels of the syllable and the prosodic word. This new research focus has been stimulated in part by developments in prosodic phonology and phonological theory that provide a framework for investigating children's early language productions in terms of a constraint-satisfaction problem, rather than as a rule-based system. This has coincided with the recent availability of new longitudinal, phonetically transcribed, computerized corpora from children learning many different languages, ideal for addressing issues of early phonological development. At the same time there has been increasing contact with other fields, including infant-speech perception, speech and hearing sciences, and computational modeling. The purpose of this chapter is to review these developments in more detail, providing the reader with a sense

The Handbook of Phonological Theory, Second Edition. Edited by John Goldsmith, Jason Riggle, and Alan C. L. Yu

of where the field began, where it is today, and some of the promising new directions for the future.

2 Early Studies of Phonological Development (1870s–1960s)

2.1 *From Descriptivist to Behaviorist Approaches to Phonological Acquisition*

Scientists and parents alike have long been interested in language development (e.g. Darwin 1877; Deville 1891). Many of these early works constituted diary studies, where the observant parent wrote down the child's words or utterances. One of the most influential in terms of phonological acquisition was Leopold's (1939–1949) study of his daughter Hildegard's language development from birth till 2 years. The fact that she was developing bilingually (English and German) made this study all the more interesting. Although there are obvious limitations to diary study methods, such as the objectivity, accuracy, and training of the observer, they have provided some of the most complete records of individual children's early phonological development. These studies typically include lists of children's early words, giving some indication of early word and syllable shapes, the syllables and segments that were omitted, and the types of segments that were modified. Such diary data are therefore ideal for investigating segmental and word-level phonological development. They are less amenable to examining the use of words in larger prosodic context, or for examining the development of morpho-phonological interactions.

The individual case study nature of these early data sources, and the lack of any overarching theoretical perspectives at the time, led to a descriptive view of the data. This changed with the rise of behaviorist perspectives after World War I, with a shift in focus to more large-scale controlled studies of language development. The most influential normative study of this period was Templin (1957), where 430 children aged 3–8 were assessed for their production and discrimination of segments, as well as receptive vocabulary and sentence production skills, laying the groundwork for future research on normative language development.

With the rise of behaviorism, Leonard Bloomfield (1933) began to see children's language development as being determined and shaped by the environment in terms of imitating what was heard in order to communicate one's thoughts and desires. He proposed that the child acquired words separately in comprehension and production, and only later brought the two together (cf. Kiparsky and Menn 1977 and Smolensky 1996b for later views). In keeping with behaviorist thoughts at the time, Skinner (1957) also had the notion that the child's ill-formed utterances would be corrected through explicit negative evidence, something that further research has shown to be rare and ineffective at modifying children's phonological or morphological representations.

2.2 *From Structuralism to Generative Approaches to Phonological Acquisition*

In the meantime, the groundwork for formal linguistics was being laid in Europe. Ferdinand de Saussure (1916) was the first to move beyond surface segmental contrasts to propose a more abstract unit, the phoneme. This was incorporated into the study of Slavic languages by the Prague Circle in the 1920s and 1930s, where Nikolai Trubetzkoy (1939) and Roman Jakobson (1941) went beyond the notion of phonemic inventories to think of language in terms of phonological systems. Roman Jakobson (1941) proposed that all children begin the process of language acquisition by producing a maximally different set of unmarked contrasts (e.g. /p/, /n/, /ɑ/). This specific proposal has never been verified at the segmental level (e.g. Velten 1943). However, the notion that phonologically unmarked structures (e.g. stop consonants, CV syllable structures) will be the first to be acquired, has generally been upheld (cf. Lindblom 1992; Locke 1983), and has had a recent resurgence of influence in the field (see Section 5.1).

Noam Chomsky's (1957) publication of *Syntactic Structures* began to develop the notion of abstract (underlying) representations in syntax. This was followed by *Aspects of the Theory of Syntax* (1965), where Chomsky argued that language was too hierarchical and noisy to be learned merely from interactions with the world. Rather, he proposed that learners were constrained by a set of universal underlying principles that guided the language learning process. This nativist (or rationalist) approach to language was in direct contrast to the previously held behaviorist approaches to language acquisition.

The structuralist focus on surface phonological feature contrasts had been an advance over previous surface segmental approaches to phonology, but it had little to say about suprasegmental phenomena such as stress. This led to the development of generative phonology in the mid 1960s, with *The Sound Pattern of English* (Chomsky and Halle 1968) providing the foundation for early generative phonological theory. As in case of syntax, it was realized that an adequate theory of phonology must characterize what a speaker (and learner) knows about the sound system of the language at an abstract level of phonological representation. This laid the groundwork for both more formal linguistic, as well as more cognitive/biological approaches to phonological acquisition.

3 The 1970s and 1980s: Generative vs. Functionalist Approaches to Phonological Acquisition

The 1970s saw the rapid rise of studies in the acquisition of phonology. Some took a generative approach, formulating a comprehensive set of phonological rules to account for children's early productions (e.g. Smith 1973). Others took a more data driven approach, explaining individual difference in terms of cognitive and biological self-organizing systems (Ferguson and Farwell 1975; Lindblom 1992;

see Vihman 1996 for a review). This section highlights the important contributions and limitations of both.

3.1 *Phonological Rules (and Constraints)*

The first comprehensive longitudinal case study of a child's phonological development came from Neil Smith's (1973) diary study of his son Amahl. This was a landmark study in terms of providing a generative, rule-based account of the child's phonological development between 2;2 and 4 years. Smith proposed that the child's productions were not a result of misperception, but rather the child's own rule-based system. Further support for this proposal was the finding of widespread across-the-board phonological processes, similar to some of those found in both synchronic and diachronic phonology. One of the interesting phonological findings was that of chain shifts (*puzzle* > *puddle* > *pickle*), analyzed further by Macken (1980). This study has had a lasting effect on the field by publishing all the data, providing a rich source of information for reanalysis as phonological theory has developed over time. A few other studies also took a comprehensive rule-based approach to understanding children's early phonologies (e.g. Ingram 1974; Macken 1987; and see Macken 1995 for a review). Although most of these examined segmental phenomena, Judith Hochberg (1988a, b) examined the acquisition of stress by Spanish-speaking children.

One of the drawbacks of some of these early studies, however, was that many of the phonological processes proposed seemed unnatural, not being attested in the phonology of any language. The tools that generative phonology had to offer at the time were inadequate for dealing with many aspects of synchronic and diachronic phonological systems, as well as for processes of acquisition. Somewhat ahead of his time, David Stampe (1969) had proposed a perspective on developing phonologies and phonological process more generally involving the notion of constraints. Also emergent at the time were proposals for autosegmental representations, initially used to help handle tonal systems (Goldsmith 1976), but eventually applied to vowel harmony and other phonological systems as well. Andrew Spencer (1986) subsequently reanalyzed Neil Smith's (1973) acquisition data, showing how a more autosegmental approach could handle certain phonological processes that were unnatural from a linear approach to segmental phonology.

Although much of the early research on the acquisition of phonology focused on segments, some European researchers began to focus on the word as an important unit in children's early phonological organization. Drawing on insights from J. R. Firth (1948), Natalie Waterson (1971, 1987) proposed that children's early phonologies could best be characterized by a holistic, non-segmental-prosodic approach. These findings were followed by proposals by George Allen and Sarah Hawkins (1978, 1980) that English-speaking children's early words tended to take the form of disyllabic trochaic feet. They observed that children's early words are often augmented (CVC > CVCV) or truncated (e.g. *banana* > ['nænə]), both processes resulting in a disyllabic trochaic foot. They further proposed that such early

word shapes might be universal, representing the default, or unmarked form of early words.

Subsequent theoretical developments at the prosody-syntax interface (Selkirk 1984) coincided with further interest in examining higher-level structures in children's early speech as well. For example, Matthei (1989) investigated across-word process in children's early speech. Consistent with Allen and Hawkins (1978, 1980), he found that some lexical items were augmented to a disyllabic trochaic foot when produced in isolation (1a–b). However, when the two are combined into a larger phonological phrase, both were phonologically reduced (1c), again yielding a binary foot.

(1)		*Child*	*Adult target*		
	(a)	['bebi]	/'bebi/	'baby'	(1;5)
	(b)	['bʊkɔ]	/'bʊk/	'book'	
	(c)	['bebʊ]	/'bebiz 'bʌk/	'baby's book'	

These findings are extremely relevant for understanding the nature of phonological constraints on the shape of children's early words, especially given more recent developments in phonological theory (see Sections 5.1 and 5.3).

Around the same time, Marlys Macken (1978, 1979) began to observe templatic types of patterns in children's early words. Although somewhat in advance of John McCarthy's (1981, 1989) work on templatic phonology, the observations were similar. In particular, Macken noted that some children went through a period of development where their early words exhibited certain distributions of consonants, such as permitting CVCV sequences to contain consonants with only [+ant, −cor] features in onset position, and only [+ant, +cor] in medial position. Thus, words such as Spanish *Fernando* were realized as [mano], and *libro* 'book' as [pito].

Researchers had long known that the phonetic transcription of speech renders only a rough approximation of what was actually said. Although it is possible to conduct broad phonemic transcription with a relatively high level of consistency, the transcription of narrow phonetic detail (e.g. stress, voicing, vowel length, intonation contours) is not as reliable. This is even more problematic in the case of child speech. This presents a problem for understanding the nature of children's early phonological representations, and how they develop. For example, English-speaking children often appear to devoice final consonants (as in German). However, Stoel-Gammon and Buder (1999) note that the reliability of transcribing voicing in child speech is notoriously poor. Macken and Barton (1980a, b) found that, despite the apparent lack of a voicing distinction on stop onsets, both English- and Spanish-speaking children make an acoustic distinction in VOT between voiced and voiceless stops. Thus, a voicing contrast appears to be distinctive in English-speaking children, despite the fact that it may be unreliably perceived as such by adults. The finding of such *covert contrasts* has important implications for our understanding of the nature of feature representations in children's early phonologies, and points to the importance of conducting acoustic analysis for determining which feature contrasts children actually make. The

extent of covert contrasts in children's early speech is not entirely known. As noted in Section 5.2, this type of research is gaining ground again, due in part to a renewed interest in gestural approaches to phonology (e.g. Browman and Goldstein 1990) and acoustic approaches to feature cues (e.g. Stevens and Keyser 2010).

Drawing on Paul Kiparsky's (1973) developments in lexical phonology (as a way to constrain the ordering of rules on phonological representations), Kiparsky and Menn (1977) proposed a two-lexicon model for phonological acquisition: a perception representation and a production representation, and a process that mapped between them (see also Menn and Matthei 1992). More recently, Paul Smolensky (1996) has argued that positing two lexicons is not necessary, showing that the same ranking of constraints can be used in both perception and production, leading to well-formed lexical representations, but constrained output forms. Interestingly, Menn (1983), drawing on earlier discussions by Charles Kisseberth (1970) involving what he called *conspiracies*, also proposed that children's early phonologies were subject to output constraints. Thus, the seeds for a constraint-based approach to the acquisition of phonology, as well as phonology more generally, had been laid.

In sum, the early formulation of generative phonology, using segments and rules, was inadequate for handling many phonological processes, both in language, and in acquisition. This led to new developments in phonological theory, exploring non-linear, prosodic approaches to certain phonological processes. As shown below, it also led to a more empiricist cognitive/biological approach to understanding phonological development.

3.2 *Cognitive and Biological Approaches to Individual Differences*

One of the problems in the field of phonological acquisition has always been the lack of data. Many studies, like Neil Smith's (1973) study of Amahl, were individual case studies of only one child, limiting the ability to generalize about phonological development both within and across languages. One of the goals of the Stanford Child Phonology Project (1968–1988) was therefore to collect longitudinal audio recordings from several (e.g. five) children from different languages.

Ferguson and colleagues (Ferguson and Farwell 1975) conducted a large, cross-linguistic study (English, French, Japanese, Swedish, Ukrainian) examining children's transition from babbling to first words (birth to 18 months). The results showed that there was no silent period between the two (in contrast with what was proposed by Jakobson 1941). A number of researchers found that there was a close relationship between the sounds and sound sequences of babbling (at 6–8 months) and those used in first words (Cruttenden 1970; Menyuk 1968; Oller *et al.* 1976). It was also found that many of children's very first words appeared to be well formed, and then went through a period of reanalysis/generalization before once again becoming more adult-like (not unlike the U-shaped curve reported for morphological over-regularization) (e.g. Moskowitz 1973). However, although the first segments produced exhibited certain language-specific tendencies,

individual differences were also found for individuals learning the same language. Some of this took the form of phonological selectivity, where some children showed unexpected phonological processes (e.g. *snow* > [nos]). Ferguson and Farwell (1975) suggested that these individual differences were evidence of a cognitive strategy to language learning, where young children were "hypothesis testers," examining the articulatory space in an effort to produce certain segmental contrasts. This led to further investigation of phonological development in terms of a cognitive processes (e.g. Macken 1978, 1979; Macken and Ferguson 1981; Macken and Ferguson 1983; Menn 1971, 1976; Vihman, Ferguson, and Elbert 1986).

Although language-specific differences were found in children's early productions, Ferguson and Farwell (1975) also suggested that there were universal phonetic tendencies that were constrained by the biology of the human vocal tract and central nervous system. This raised questions about children's perceptual abilities and the relationship this had to their early language production. Locke (1983) proposed that biological models of language development might be more explanatory, more effectively handling the issue of continuity in language development. Kent (1984) was one of the first to think of early phonologies as self-organizing systems, where rhythmic structures were seen as a natural biological phenomenon that could account for some of Allen and Hawkins's (1979, 1980) observations about the prominence of trochaic feet in early grammars. MacNeilage and Davis (1990) further proposed that the observed prominence of CV syllable structures in children's early productions, as well as C-V interactions (e.g. high vowels more likely to follow coronal consonants, etc.), could best be understood in terms of constraints on mandibular oscillation. Others have incorporated ideas from dynamical systems theory, proposing that phonological development can best be explained with reference to increasing complexity of organization, or pattern formation adaptation, in terms of exploration of the sound space (Thelen 1981, 1989, 1991; Vihman 1993).

Björn Lindblom's (1983, 1992) concern with the learner's perceptual capacities, and the connection between perception and production, was consistent with his view that there was a tension in speech production between *information transfer* and *gestural economy*. That is, speakers try not to use more articulatory effort than needed to convey the intended information. The most stable representation for maximal clarity of acoustic cues is a series of CV sequences, which is also the most frequent syllabic sequence in children's canonical babbling and early speech (MacNeilage 1980; MacNeilage and Davis 1990). This may be a consequence of early lack of control of coarticulation, leading to the pervasive production of consonant harmony (Vihman 1978) and reduplicative forms in young children's early speech (e.g. Waterson 1971 *biscuit* [be:be:]). Recent developments in articulatory phonology (Browman and Goldstein 1990) have renewed interest in the role of articulatory gestures and their role in understanding phonological systems. As discussed in Section 5, these theoretical developments, as well as new (ultrasound) technology, now make it possible to explore the nature of articulatory constraints on children's early productions, and their relationship to phonological representations, in a non-invasive fashion.

In sum, the Stanford Child Phonology Project served as a major stimulus for the study of phonological acquisition beginning in the 1970s. Two conference proceedings volumes (Ferguson, Menn, and Stoel-Gammon 1992; Yeni-Komshian, Kavanagh, and Ferguson 1980) presented the state of the art at the time, and a reference point for researchers in a field that began to rapidly change in the 1990s. The Stanford studies were largely descriptive, providing for the first time developmental data from several children from five different languages from babbling until they had a vocabulary of approximately 50 words (around 1;6). However, it is around this point in development that evidence for children's phonological representations begins to be more robustly evidenced. Thus, although these studies provided ample evidence of early segmental development, and some word development, few testable predictions can be made about children's developing phonological representations using this data. Rather, what is needed is data spanning the time period from the onset of first words (around 1 year) until 2 or 2;6 years – data that, until recently, has been scarce. Also missing during this period was a theoretical framework from which to understand the nature of developing phonological representations. As discussed in the next section, both theory and data began to arrive in the 1990s, stimulating the field of phonology and phonological acquisition in new ways.

4 From the 1990s to 2007: Phonological Constraints

The 1990s had a significant impact on both the field of theoretical phonology and phonological acquisition. This was largely due to the stimulus provided by Optimality Theory (Prince and Smolensky 2004), which energized the entire field of phonological inquiry. However, the 1990s also saw the beginnings of more interdisciplinary research in phonological acquisition. One of the venues that provided a more theoretical stimulus to the field was the University of British Columbia's International Conference on Phonological Acquisition (Bernhardt, Gilbert, and Ingram 1996). This event brought together a wide range of European and North American speech-language and phonological researchers working on a wide variety of languages, stimulating new and continuing collaborations in the field. During this time researchers of language acquisition and infant-speech perception also began to interact, stimulated by an interdisciplinary conference at Brown University (Morgan and Demuth 1996). This was followed by a similarly inspired conference in Berlin, resulting in the volume by Jürgen Weissenborn and Barbara Höhle (2001). Both brought together researchers working on the perception and production of different languages, as well as more computationally oriented researchers, laying the groundwork for future collaborative endeavors. These interactions were reinforced by the broadening of the scope of papers accepted at the yearly Boston University Conference on Language Development to include aspects of infant-speech perception and computational approaches to learnability in addition to the more traditional theoretical linguistic treatments of language acquisition.

The 1990s therefore had a significant impact on the field of phonological acquisition, and this influence is ongoing today, with a growing international community of researchers working on the acquisition of phonology. This has come about, in part, through the development of a clearer set of predictions about the course of acquisition and how this might be realized in languages with different phonological structures. But it has only been possible to address these issues with the increased availability of longitudinal computerized corpora from more children between the ages of 1 and 3. These developments have led to a number of special conference sessions and special thematic volumes devoted to the acquisition of phonology, again providing an important reference for the training of future researchers. For example, Goad and Rose (2003) focused on segmental-prosodic interactions, with several papers discussing constraints on cluster reduction. René Kager, Joe Pater, and Wim Zonneveld (2004) included papers examining a variety of issues in phonological acquisition from an optimality-theoretic perspective. The papers in Demuth (2006) explore the impact of frequency effects on understanding cross-linguistic differences in prosodic word shape. All provide insight into the acquisition process as one of constraint satisfaction.

4.1 *The Emergence of the Unmarked in Early Phonologies*

As mentioned in Section 3, acquisition researchers had already experienced the limitations of rule-based, segmental accounts of children's early productions (Smith 1973), and had begun to explore other approaches to understanding the nature of early phonological systems. This continued in the 1990s. For example, Demuth (1993) used an autosegmental approach to the acquisition of Bantu tonal systems, showing that 2-year-old Sesotho speakers had no problem with lexical tone, but only acquired grammatical melodies that interacted with tone sandhi and OCP effects by the age of 3 (see Yavas 1994 and Archibald 1995 for other non-linear treatments of both first- and second-language acquisition). The development of a constraint-based approach to phonology (Prince and Smolensky 2004) provided further tools for exploring the nature of phonological acquisition.

As previously mentioned, one of the challenges for the field of phonological acquisition has always been the lack of data. It is relatively easy to conduct phonological and syntactic experiments with older children, and corpora available on the CHILDES database (MacWhinney 2000) include a rich source of information about children's syntactic structures. However, in 1990 there was little in the way of phonetically transcribed acquisition data available in the public domain. The publication of Paula Fikkert's (1994) thesis on the phonological development of 12 Dutch-speaking children, which took a parameter-setting approach to the acquisition of stress (Dresher and Kaye 1990), was therefore a significant event given the wealth of data it contained.

At the same time, Jane Fee (1995), and Katherine Demuth and Jane Fee were examining English-speaking children's early word shapes from a more prosodic perspective, trying to provide a unified explanation for both weak initial-syllable truncation and reduplication/vowel epenthesis. Drawing on developments in

prosodic phonology and morphology (Nespor and Vogel 1986; Selkirk 1984, 1996), they proposed that children's early productions are governed by highly-ranked No-Coda constraints, as well as constraints against initial unstressed syllables. They also suggested that children's early words were actually "minimal words," and that children's grammars provided support for the *emergence of the unmarked* (McCarthy and Prince 1994b), providing the first optimality-theoretic (OT) analysis of children's prosodic words; Demuth 1995). Demuth (1996) then reanalyzed Fikkert's Dutch data, showing that, like the English data, it could be handled from a constraint-based perspective, and that perhaps children learning all languages would exhibit a similar minimal word stage of early development. Gnanadesikan (2004) similarly proposed that the notion of the emergence of the unmarked could help account for children's early onset reductions at the level of the syllable. Pater (1997) then integrated these two proposals, showing that the children's early word truncations could be understood in terms of both higher-level prosodic constraints that also obeyed markedness constraints at the level of the syllable. Thus, the truncation of *banana* to ['bænə] preserved an obstruent at the beginning of the syllable, the unmarked option for onsets.

Since that time, there have been numerous OT analyses of children's developing phonologies, and even those that do not provide a formal OT analysis are often inspired by the notion of constraints. Some of this was anticipated by phonologists such as Stampe (1969), who, like Kisseberth (1970) and Paradis (1988), saw phonological systems as a constraint-satisfaction problem rather than as a set of phonological rules. This perspective has provided the field of phonological acquisition with an extremely useful framework from which to explore the nature of children's developing prosodic phonologies. We provide the highlights of some of these studies below.

4.2 *The Acquisition of Syllable Structure*

Although Erik Fudge (1969) had pointed to the importance of the syllable as a unit of phonological analysis, this was a relatively neglected area of research in the generative program until the work of G. N. Clements and Samuel Keyser (1983) (see Chapter 5, this volume). The development of notions relating to the sonority hierarchy and the sonority sequencing principle (e.g. Clements 1990; Ladefoged 1993) set the stage for examining further phonotactic restrictions on syllable structure acquisition. For example, Kehoe and Stoel-Gammon (2001) found that alveolar stops were the first coda consonants to be acquired in English. This was of significant interest since sonorants are typically assumed to be the unmarked form for coda consonants crosslinguistically, exhibiting less of a sonority rise between nucleus and coda than from onset to nucleus. However, Stites *et al.* (2004) showed that alveolar stops are the most frequent coda consonants in English, and confirmed that children tend to acquire these first, rather than the less frequent, phonologically unmarked sonorant codas. Thus, although frequency and markedness typically pattern together, most children may show a preference for frequency over markedness effects in their early productions, all else being equal. This raises

questions about the notion of markedness as a whole, and its relationship to frequency for learners of a particular language. It also raises the question of which linguistic units learners are using for calculating frequency. For example, Zamuner *et al.* (2004) show that coda consonant production is a function of neighborhood density. That is, it is the frequency of the rhyme+coda, rather than simply the coda consonant itself, that is the best predictor of accuracy in coda consonant production, at least English. On the other hand, /ʁ/ is one of the most frequent consonants in French, yet several studies have found that at least some French-speaking children have persistent problems with the production of /ʁ/ (e.g. Demuth and McCullough 2009a; dos Santos 2007; Rose 2000). This may be due to articulatory problems with this uvular fricative, or due to its variable realization in the input children hear.

Developments in feature geometry (e.g. Clements 1985) and gestural approaches to phonology (Browman and Goldstein 1990) also provided the theoretical background for examining consonant-vowel interactions within the syllable. For example, Claartje Levelt (1995) found that Dutch-learning children were much more likely to produce onset consonants that shared the features of the following vowel. Thus, the consonants preceding back rounded vowels tend to be labials, and those preceding high vowels tend to be coronals. This raises the question of when and how learners begin to be able to represent cues to feature contrasts in an adult-like fashion. This will be explored further in Section 5.

Research on the structure of the syllable has also given rise to many studies on the acquisition of more complex syllable structures. Much of the early work on consonant clusters was carried out with children who exhibited phonological delay, raising questions about the representation of clusters in children's early phonologies more generally (e.g. Chin and Dinnsen 1992; Gierut 1999; see Bernhardt and Stemberger 1998 for a review). Some of this research focused on the factors governing cluster reduction, with different proposals as to the constraints involved. Following Pater (1997), many researchers have proposed that children will typically preserve the least marked onset, i.e. the least sonorant segment of the cluster (e.g. Barlow 1997; Diane Ohala 1996, 1999). Thus, in the case of the word *stop*, the obstruent /t/ would be preserved, but in the case of the word *sleep*, the /s/ would be preserved.

However, others have noted the limitations of the sonority account (e.g. Barlow 1997, 2001; Goad and Rose 2004; Pater and Barlow 2003). Goad and Rose (2004) proposed that children preserve the consonant that is the head of the syllable. Thus, in an obstruent + liquid cluster such as *plate*, the obstruent /p/ is the head of the syllable and will be preserved. In contrast, the /s/ in s-clusters is syllabified as an adjunct, leaving the /l/ as the preserved head of the syllable in a word like *slate*. However, Pater and Barlow (2003) show that one child simplified the onset of *sneeze* to /n/, whereas the onset to *sleep* was simplified to /s/. Jongstra (2003) therefore proposed that when the sonority distance is close, the segment contiguous with the nucleus will be preserved, whereas when the sonority distance is sufficiently far, the least sonorous segment will be preserved. A recent study of cluster simplification calls all the above into question, noting that features from

both consonants often remain in cluster reduction (Kirk 2008). Interestingly, the majority of studies of consonant cluster reduction have examined the acquisition of Germanic languages. Little is known about consonant cluster reduction processes in other languages.

All of the above mentioned studies examined onset clusters. There has been little research on the acquisition of word-final clusters. One might expect these to be acquired later due to the fact that codas are typically considered more marked than onsets. However, Lleó and Prinz (1996), in a longitudinal study of five German-speaking 1–2 year-olds, found that word-final clusters were acquired several months earlier than word-initial clusters. Furthermore, Levelt *et al.* (2000) found that nine of the children in the Dutch CLPF corpus first acquired clusters word-finally, whereas only three of the children first acquired clusters word-initially. They suggest that this variability in Dutch patterns of syllable structure development is due to the fact that the frequency of consonant clusters in both positions is approximately the same. Controlling for segments, Kirk and Demuth (2005) found that English-speaking 2-year-olds were also more accurate at producing word-final as opposed to word-initial consonant clusters. Overall, coda clusters are more frequent than onset clusters in English. However, the children also exhibited better production of final nasal+s and stop+s clusters than final nasal+stop and s+stop clusters. Furthermore, children often metathesize the s+stop clusters (*wasp* > *waps*), suggesting that frequency or articulatory factors may be involved. Note also that the most accurately produced clusters are those that typically occur with morphologically complex forms, suggesting that morphology may provide a perceptual or production advantage for some consonant clusters.

To control for the possible effects of frequency and morphology, Demuth and Kehoe (2006) conducted an elicited production study to examine the acquisition of consonant clusters in French. The found that 2 year-olds were more accurate at producing onset rather than word-final clusters, a finding confirmed in a subsequent longitudinal study (Demuth and McCullough 2009a). Some researchers have proposed that some word-final consonants in French (and other languages) prosodify as onsets to empty-headed syllables (e.g. *partir* 'to leave' /paʁ.ti.ʁØ/) (Charette 1991). It is possible that this structure is more marked, and therefore acquired later (though Goad and Brannen 2003 claim that such structures are universal at early stages of acquisition). Rose (2000) noted, however, that one child from his longitudinal study of two children learning Canadian French had acquired all but /ʁ/ in word-final position, but had /ʁ/ as a coda word-internally. He therefore proposed that this child had a coda representation for /ʁ/ in all positions. However, researchers have also reported that the acoustic and articulatory characteristics of French /ʁ/ are extremely variable, both between speakers and within the speech of individual speakers (see Demuth and McCullough 2009a for review). Little is known about the acquisition of segments that are variably realized in the input, or where the syllabic representation is potentially ambiguous (though see discussion of the acquisition of branching onsets and rising diphthongs in Rose (2000) and Kehoe *et al.* 2008).

4.3 *The Acquisition of Prosodic Word Structure*

The early work on the acquisition of prosodic word structure (Demuth 1995; Pater 1997) suggested that children had an early awareness of word-minimality effects, and that this could be captured in terms of constraint interactions. Further support for this proposal came from Mitsuhiko Ota (1999). Using acoustic analysis, he showed that young Japanese-speakers who cannot yet produce coda consonants exhibit compensatory lengthening of the vowel, thereby preserving bimoraic (and minimal word) structure. In many respects, this study was ahead of its time, providing acoustic/phonetic evidence to support the theoretical claims made.

One of the issues raised by the Word Minimality Hypothesis was what predictions it would make for the acquisition of a language such as French, where a large portion of the lexicon violates word-minimality. Demuth and Johnson (2003) examined this issue in longitudinal data from one French-speaking child. They found that the child's earliest words (1;3–1;5) were all target or reduplicated CVCV forms. As in other languages, her early grammar showed a highly-ranked constraint against word-final (coda) consonants, resulting in either reduplicated CVCV repairs, or truncated CV outputs. Thus, for certain CVC target words she produced subminimal, monomoraic words. More striking, however, was the reduction of disyllabic CVCV words to monosyllabic CV form. Further analysis showed that segmental constraints against fricatives, velar stops, and clusters were more highly ranked than faithfulness to syllable preservation and/or word minimality (see dos Santos 2007 for similar observations from another child who does have velar consonants). French is also a language that permits subminimal, monomoraic CV lexical items, and these constitute approximately 20% of all words French-speaking children hear (Demuth and Johnson 2003). In keeping with similar proposals by Levelt *et al.* (2000), the authors suggest that learners are sensitive to the high-frequency phonological structures of the ambient language, and adjust their grammars (constraint ranking) accordingly. This can be understood in terms of trying to be as faithful as possible to the input forms, thereby minimizing constraint violations. Note that such a perspective on the development of early grammars minimizes the role of universal markedness. Rather, higher-frequency phonological forms become the "unmarked" structures on a language-specific basis.

On the other hand, Goad and Buckley (2006) proposed that French learners do show early word-minimality effects. They report that a Canadian child showed compensatory vowel lengthening when the word-final consonant was missing, though no supporting acoustic analysis was provided. However, further analysis of spontaneous productions from two children from France showed no systematic lengthening of the vowel when the word-final consonant was missing (Demuth and Tremblay 2008). The number of subjects examined in all these studies was small, suggesting that further study with more children at the early stages of acquisition (1–2 years) is needed to resolve this issue.

Demuth *et al.* (2006) returned to the issue of word-minimality in English, drawing on new data from four children between the ages of 1–3. Although some

children showed apparent compensatory vowel lengthening (and one child showed an early period of epenthesis, where CVC words surfaced as CVCV), this occurred on both monosyllabic and disyllabic words, and on both long/tense as well as short/lax vowels. If learners were using compensatory lengthening to preserve word-minimality, one would expect this to be restricted to monosyllabic targets with short/lax vowels, the context where a second mora of structure is required to preserve a bimoraic foot, or minimal word. Further acoustic analysis of three children's compensatory processes found that two of the children exhibited across-the-board compensatory lengthening for missing codas, whereas only one (older) child showed compensatory lengthening only for target words with a short/lax vowel (Song and Demuth 2008). These findings suggest that children may initially exhibit compensatory lengthening for omitted coda segments in English, and only later (around the age of two) come to realize that languages like English observe word-minimality constraints.

Note that the implications of the English findings on compensatory vowel lengthening contrast with those of Mitsuhiko Ota (1999), where he proposed that early vowel lengthening to compensate for missing codas in Japanese provided support for an early awareness of moraic structure. There are three possible explanations for this. First, Japanese is a mora-timed language, and children may become more aware of moraic structure and its consequences for prosodic word structure earlier in such a language. Second, although Japanese, like French, permits subminimal words, Ota reports that Japanese parents, in speaking to their young children, generally augment subminimal forms. This raises the issue of the nature of the input children hear (child-directed speech), and the effects this may have on children's developing phonologies. Finally, it is possible that both Japanese- and English-speakers compensate for missing segments at the ends of words. However, since coda consonants are always moraic in Japanese, it is difficult to determine if compensatory lengthening is due to segmental versus moraic factors. This is obviously an area for further cross-linguistic research.

Roark and Demuth (2000) proposed that the frequency distribution of syllable and prosodic word structures in the input children hear could influence the types of prosodic word structures children used in their early utterances. In a corpus analysis of child-directed speech they showed that the token frequencies of word shapes in English and Spanish are significantly different. In particular, the majority of words in English are monosyllabic, and there are very few trisyllabic words. This contrasts with Spanish, which has many more trisyllabic and tetrasyllabic words. They suggested that this different distribution of word shapes could account for English-speaking children's tendency to truncate words like *banana* until around the age of 2;6 (Pater 1997). In contrast, Spanish-speaking children appear to permit prosodic words of larger structure earlier than English-speaking children. Several of the papers in Demuth (2006) pursue this issue further, bringing a much-needed cross-linguistic perspective on these issues. Vigário, Freitas, and Frota (2006) and Lleó (2006), analyzing acquisition data from European Portuguese and Spanish respectively, found that the relative frequency of word shapes in the input helps explain the truncation patterns. This is consistent with the findings from English,

Dutch, and French mentioned above. However, Prieto (2006), in a comparative analysis of the acquisition of Spanish and Catalan, suggested that the relative frequency of foot shape, rather than prosodic word shape, helps explain why Catalan learners (but not Spanish learners) pass through a stage of development where they truncate disyllabic S(w) prosodic words. Finally, Ota (2006) suggested that lexical frequency, rather than prosodic word shape, best accounted for the few cases of truncation found in children's Japanese. Thus, frequency effects at various levels of structure may help determine constraint rankings in the grammars of children learning different languages, resulting in different truncation patterns in early prosodic word development.

Importantly, these patterns of truncation appear to be due to phonological, not perceptual or articulatory constraints. Interestingly, Carter and Gerken (2004) found that, when children omitted the initial unstressed syllable of a three-syllable word following a verb, they left a prosodic trace of the missing syllable, which is realized as a silent interval. This suggests that, at least in some cases of syllable omission, children have in some sense planned for the syllable, even though no segmental content is realized. Such findings provide further evidence of covert contrasts in children's early speech that are often missed in traditional phonetic transcription. These findings raise questions about the extent to which other apparent omissions in early child speech may in fact be prosodically, gesturally, or acoustically realized at some level of analysis. They also point to the need for a developmental model of speech planning and production. Both would provide a better understanding of the phonological representations and phonological constraints that govern children's early phonological grammars.

4.4 *Acquisition at the Phonology/Morphology Interface*

Drawing on insights from the Prosodic Hierarchy, and proposals for the prosodification of grammatical morphemes, researchers in the 1990s also began to examine the nature of prosodic constraints on children's realization of grammatical morphemes. Early research on southern Bantu languages (which exhibit penultimate lengthening at the end of a phonological phrase) had noted that children tended to produce noun class prefixes with monosyllabic stems, but omit them with disyllabic stems (e.g. Connelly 1984). Demuth (1994) proposed a phonological explanation of this issue, observing that children tended to produce noun class prefixes when these constituted part of a disyllabic foot (*mo-tho* 'person'), but were more likely to omit them when they were unfooted (*mo-sadi* > [sadi] 'woman'), much like the truncation of prosodically similar English monomorphemic words (e.g. *banana* > ['nænə]). This phenomenon begins to disappear around the age of 2;3, once children's prosodic phonologies develop further (Demuth and Ellis 2009). LouAnn Gerken and colleagues (Gerken 1994; Gerken and McIntosh 1993) also noted that English learners were more likely to produce grammatical morphemes such as pronouns and determiners when these could be prosodified as part of a foot. In the meantime, Selkirk (1996) proposed a set of constraints on prosodic well-formedness, where prosodic structures immediately dominated by the next

higher level of structure within the Prosodic Hierarchy conformed to the constraint on exhaustive prosodic parsing, or Exhaustivity. Thus, prosodic clitics that could not be prosodified as part of a foot would violate Exhaustivity. Gerken (1996) capitalized on this development, proposing that children would be more likely to include grammatical morphemes in early speech if these could be prosodified as part of a foot, thereby conforming with the constraint on Exhaustivity. However, unfooted grammatical morphemes would need to be prosodified at a higher level of structure (e.g. the phonological phrase), and would therefore violate the Exhaustivity constraint. Thus, children's variable omission of grammatical function items could be understood in terms of markedness, where those determiners that incurred the fewest constraint violations would appear earlier in children's productions.

Lleó (1996) had long noted that Spanish-speaking children (unlike German-speaking children) exhibit the use of proto-determiners from the beginning of their speech. This was explained in terms of the high frequency of Spanish three-syllable words, which provide children below the age of 2 with a prosodic window large enough to permit the prosodic licensing of determiners (Demuth 2001; Lleó 2001; Lleó and Demuth 1999). Support for this hypothesis came from the fact that three-syllable words that were truncated to two syllables are nonetheless accompanied by a (proto)determiner (e.g. *la muñeca* 'the doll' > [a'meka] (Demuth 2001, Demuth *et al.* in press)). This suggests that Spanish-speaking children have an early three-syllable prosodic word window that can be adapted to permit inclusion of a determiner plus a following foot.

Research on other languages shows a similar tendency for children to produce prosodically licensed grammatical morphemes earlier than those that are not, thereby accounting for some of the variable production of grammatical function items in children's early speech. For example, Demuth and Tremblay (2008), in a study of two French-speaking children, showed that determiners begin to robustly appear with monosyllabic words around 1;10 years, but only begin to be consistently used with disyllabic words a few months later. The authors suggest that the early determiners are prosodified as part of an iambic foot. Only once children begin to represent higher levels of prosodic structure do they begin to include determiners with longer words, showing evidence of having acquired adult-like prosodic structure, where determiners are prosodified at the level of the phonological phrase.

Demuth and McCullough (2009b) found a comparable pattern of longitudinal development for the acquisition of articles by five English-speaking children. Similar to the French findings, and consistent with the experimental findings from Gerken (1996), they show that four of the children had significantly higher use of articles when these could be prosodified as part of a foot with the preceding word. In contrast, the children tended to omit the articles that remained unfooted (i.e. those prosodified at the level of the phonological phrase) (e.g. *Tom [hit the]*$_{FT}$ *ball* vs. *Tom [patted]*$_{FT}$ *(the) ball*). Furthermore, this pattern persisted for 4–5 months, disappearing as the children reached the age or 2–2;6 years. This is about the same time that children begin to more reliably produce the initial unstressed syllables of lexical items like *banana* (cf. Pater 1997).

These findings suggest that children's acquisition of grammatical morphemes is closely tied to the development of prosodic representations. Given that many grammatical morphemes are unstressed prosodic clitics, their acquisition is dependent on the development of higher-level prosodic structures. The Prosodic Licensing Hypothesis therefore provides a framework for exploring the development of higher-level prosodic representations, and how this changes over time. It also provides a principled means for making predictions about the course of grammatical morpheme development within and across languages. As shown in the case of Spanish determiner acquisition, the development of these constraints on prosodic structure is also closely tied to the prosodic properties of the lexicon, though the exact relationship between the two is not yet clear.

The prosodic licensing of grammatical morphemes appears to occur at the level of the syllable as well (e.g. Stemberger and Bernhardt 1997). Recent research has found that there are phonotactic and positional effects on the acquisition of English third-person *-s*. That is, children are much more likely to produce this grammatical morpheme when it forms a simple coda than when it forms part of a consonant cluster (e.g. *sees* vs. *hits*), and when it appears utterance finally compared to utterance medially (Song *et al.* 2009). These findings suggest that there is still much to be discovered about the phonology-syntax interface in children's developing grammars, where constraints on prosodic representations may account for much of the variable production of grammatical morphemes.

4.5 *Featural Underspecification and Phonological Processes*

With the development of approaches to feature underspecification (Archangeli 1988; see Steriade 1995 for a review) came a renewed interest in trying to understand segmental phenomena in children's early grammars. For example, Stemberger and Stoel-Gammon (1991) and Stoel-Gammon and Stemberger (1994) proposed that consonant harmony processes, which are common in early English, could be understood in terms of underspecification theory. Goad (1997) took this further, proposing that consonant harmony in children's early grammars could be characterized in terms of the relative ranking of parse/link place features and those that align place features at the left edge of the harmonic domain (see dos Santos 2007 for discussion of similar phenomena in French). Others explored which theory of underspecification (radical, contrastive) could best account for children's phonemic inventory, and how this changed over time (e.g. Gierut 1996). Morrisette *et al.* (2003) explored issues of markedness and the representation of place features, making the prediction that if the child's system represented dorsal place distinctions, it would also represent coronal place. Furthermore, they proposed that coronals would be expected to replace dorsals, but not vice versa. Interestingly, this prediction is not upheld in the cases of consonant harmony in one French-speaking child, where dorsals regularly replace coronal consonants that are not in the head of the foot (dos Santos 2007).

4.6 *Frequency Effects and Phonological Acquisition*

Every since Noam Chomsky (1965) defined the field of formal linguistics as constituting the knowledge of grammar (linguistic competence, or I-language), issues of language performance (language use, E-language) have been largely ignored, or relegated to the field of sociolinguistics. However, researchers have long been aware that issues of lexical frequency, for example, play a large role in understanding aspects of grammaticalization and historical change, many of which involve phonological and morphophonological processes (e.g. Bybee 2001, 2007). Psycholinguists have also long known that adults appear to have encoded not only information about the phonology, morphology, syntax, and semantics of lexical items, but also the likelihood that a given lexical item will appear in a given grammatical context (e.g. MacDonald *et al.* 1994). Furthermore, this notion of *predictability* is encoded in how we speak, with high-frequency and/or predictable information typically being phonologically (or at least phonetically), reduced (e.g. Aylett and Turk 2004; Lindblom 1983). That is, the speaker appears to have a model of the listener in mind, and will phonetically reduce redundant or less important information when speaking.

Recent research on infant-speech perception has shown that infants are extremely sensitive to the frequency of segments and prosodic structures in the primary language to which they are exposed (e.g. Anderson *et al.* 2003). Thus, despite the fact that there is some noise in the signal, language learners appear to be capable of extracting the information needed to create both a phonology and a lexicon (see Saffran *et al.* 1996) for discussion of whether these abilities are domain specific, or more general). It has also long been known that 3–5-year-old children's morphophonological representation of familiar, high-frequency words is more robust in both perception and production than that of novel and low-frequency words (Edwards *et al.* 2004). More recently, scholars have begun to examine more systematically the effects of frequency on the acquisition of syllable and prosodic word structures (e.g. Zamuner *et al.* 2004).

Recall that Roark and Demuth (2000), in a corpus analysis of child-directed speech, found that English had a much higher frequency of syllable final (coda) consonants than did Spanish (60% vs. 25%). These frequencies closely matched those suggested by Pierre Delattre (1965) for adult speech. They used this to help explain the earlier acquisition of coda consonants in English relative to Spanish. In a similar vein, Levelt *et al.* (2000) found that the frequency of different syllable structure shapes (e.g. CVC, etc.) in Dutch corresponded closely to the order in which these were acquired. Importantly, they also showed that the fit was much better using a corpus of child-directed speech rather than adult-directed speech, indicating that the course of acquisition adheres closely to the statistics of the lexicon children typically hear (cf. Ota 2006). Furthermore, they found that, when the frequency of syllable structures was the same (CCVC, CVCC) nine of the children acquired the less marked complexity in the onset first, and only three acquired the more marked complexity in the coda first.

These results, showing that language learners have an early sensitivity to the frequency of different phonological and prosodic structures, is interesting in light of proposals that children would first acquire crosslinguistically unmarked structures (e.g. Demuth 1995; Gnanadesikan 2004; Jakobson 1948; Pater 1997). Although unmarked structures are often extremely frequent, this is not always the case. For example, given the restriction found in many languages where only sonorant consonants are permitted in the coda (e.g. Japanese), it is often assumed that sonorants are the unmarked form for coda consonants (cf. Clements and Keyser 1983). However, in languages like English, the most frequent coda consonants are stops, the most frequent of these being /t/. That is, the highest frequency coda consonants in English are the most marked crosslinguistically, raising questions about the order of acquisition of coda consonants in this language. To investigate this issue, Stites *et al.* (2004) conducted a longitudinal study of three children, investigating which coda consonants were the first to appear. They found that two of the children showed the frequency pattern, whereas only one showed the markedness pattern. Cross-sectional findings by Kehoe and Stoel-Gammon (2001) confirm that /t/ is the first coda consonant to appear. Thus, it appears that, for most children, robust frequency effects will typically override more cross-linguistic (i.e. markedness) tendencies, at least in some prosodic domains.

One of the challenges to any study of frequency effects is to determine what to count. The results mentioned above indicate that language learners may be keeping statistics over any number of different types of linguistic units simultaneously. In the case of phonology, this might include every level of the Prosodic Hierarchy, and segmental interactions therein. For example, much of the research on lexical acquisition finds that children's accuracy in the production of lexical items is closely related to neighborhood density (Edwards, Beckman, and Munson 2004; Storkel 2004). Furthermore, the acquisition of coda consonants appears to be closely linked not only to coda consonant frequency, but also to neighborhood density within the entire rhyme (Zamuner *et al.* 2004). Thus, some of the variability found in the acquisition of syllable structures, as well as words and morphemes, may be explained by the frequency with which these are segmentally and phonotactically represented in the lexicon.

Issues of lexical frequency, as well as acoustic and articulatory factors, may account for some of the variable production of coda consonants in different prosodic contexts. For example, Demuth *et al.* (2006) showed that English-speaking children acquire word-final coda consonants earlier in monosyllabic as compared with disyllabic words. In an experimental study with novel disyllabic words, Kirk and Demuth (2006) showed that children were more likely to produce the same coda consonant when it occurred either in a stressed syllable or at the end of the word. They suggest that this is due to the longer duration found in both positions, providing increased time for the inclusion of another articulatory gesture. Thus, frequency effects appear to help explain some of the variance in phonological acquisition within a certain class of prosodic structures. But across prosodic contexts, either within or across languages, other contextual and/or gestural planning phenomena may better account for some of the variable production found. Thus,

although the frequency of coda consonants is much higher in English compared to Spanish, the earlier acquisition of coda consonants in English may also be due to the fact that most English lexical items used in everyday speech are stressed monosyllables. This contrasts with Spanish, where coda consonants often fall on unstressed syllables and/or at the ends of polysyllabic words (e.g. *escaleras* /es̲.ka.'le.ras̲/ 'stairs'). Thus, in addition to frequency effects, prosodic factors such as position within the word and phrase (Hsieh, Leonard, and Swanson 1999), as well as stress, may also play an important role in determining the nature of children's early syllable, word and morpheme productions. Such issues are not currently incorporated into models of early acquisition, in part due to the focus on linguistic competence rather than on performance factors. However, most of the child language data come from performance of some kind, be it perception, comprehension, or production (though often not recognized, this is true of adult linguistic data as well). By controlling for the factors outlined above, it may be possible to better understand the nature of children's knowledge of phonology (linguistic competence), and when various phonological structures have been acquired.

5 The Future of Phonological Acquisition

The field of phonological acquisition has grown significantly since the early 1990s. This has been largely due to theoretical developments in phonology, combined with the new availability of phonologically transcribed longitudinal data from several children in different languages. New developments are also now taking place in experimental methods, acoustic analysis, and computational modeling, pointing to a vibrant future for the field. This section highlights some of these developments and their theoretical import.

5.1 Constraint-based Approaches

The field of phonological acquisition has been significantly influenced by the development of constraint-based approaches to the study of phonological systems (e.g. Prince and Smolensky 2004). This has provided a framework for investigating interactions between different types of constraints in the developing system, and for viewing phonological acquisition as a constraint-satisfaction problem. That is, given the limited processing/production capacities of a 1–2-year-old, how does the child make himself or herself understood? Viewed as a constraint-satisfaction issue, it is possible to see that there are several different solutions to a given phonological problem, giving rise to certain types of well-attested individual variation. This approach therefore provides a framework for understanding individual variation in terms of differently ranked (or weighted) constraints, much in the way that dialectal variation occurs in closely related languages. Thus, although Optimality Theory and related approaches provide little in the way of predictions about what is possible or impossible in early acquisition (this

is presumably governed by constraints on universal grammar as well as the developing physiology of the child), it does provide a framework for exploring the nature of constraint interactions for a given child learning a given language. What determines the particular ranking of constraints, and how this changes over time, will presumably be determined, in part, by the frequency of different constraint violations in conjunction with general markedness factors. Thus, given an initially highly-ranked constraint against coda consonants (No-Coda), the learner of a language with many codas (e.g. German or English), will typically demote this constraint faster (to avoid massive constraint violations) than the learner of a language with fewer codas (e.g. Spanish) – all else being equal (e.g. Boersma and Levelt 2000). Furthermore, some of the variability found within a given child, for the same target word in the same sentential context, can be handled in much the same way that variability has been handled in adult phonological systems (e.g. Anttila 2002; Nagy and Reynolds 1997), that is, in terms of floating or overlapping constraints (Boersma and Levelt 2000; Demuth 1997).

As in the field of phonology itself, each of these issues was intractable within the framework of a rule-based system. A constraint-based perspective therefore provides a means for better describing what is happening during the acquisition process. This in turn lays the groundwork for developing an explanation for the process, and making predictions about how a given phonological phenomenon will be acquired in other domains or other languages. To address these issues we will need access to new data sources, something that has already begun to develop (see Section 5.3).

5.2 *Articulatory/Acoustic Approaches*

At the same time that constraint-based perspectives on phonology began to grow, articulatory approaches to phonology also began to develop (Browman and Goldstein 1990). If feature contrasts are the outcome of articulatory gestures, then surely this has implications for understanding the nature of early phonological acquisition as well. Although this perspective on language acquisition has been slower to develop, it is beginning to be investigated more seriously in cases of disordered speech. For example, using ultrasound technology, Bernhardt *et al.* (2005) have found that English-speakers who exhibited persistent problems with the production of /ɹ/ were using only one of the articulatory gestures needed to produce this segment. Once they were given intervention on the appropriate gesture to use, they quickly began to pronounce /ɹ/ as appropriate to a North American west coast dialect. Thus, some of the constraints on children's acquisition of phonology may be articulatory, especially in cases where multiple articulatory gestures are required (e.g. liquids, clusters, affricates). This is obviously an area for future research.

Given the complexities of articulation, there may also be acoustic and/or articulatory evidence that children are actually approximating certain feature contrasts, despite the fact that these are often not perceived as such by the listener/transcriber. In the past few years there has been renewed interest in investigating

such covert contrasts (e.g. Scobbie *et al.* 2000), and we expect such lines of inquiry to continue, providing a richer set of acoustic evidence for children's developing phonological representations. Critical to such an endeavor would be the investigation of how children represent certain acoustic landmarks in their early speech productions, and how they enhance certain gestures or acoustic features to ensure that certain feature contrasts are clear (e.g. Keyser and Stevens 2006; Stevens and Keyser 2010). Stoel-Gammon and Buder (2002) showed that most English-speaking children control extrinsic vowel lengthening before voiced/voiceless consonants by the age of 2. However, Demuth *et al.* (2006) also suggest that those children who exhibit a period of word-final vowel epenthesis on CVC targets, may do so to ensure that the voicing cues to the final consonants are clear (see Section 4.3). Further research has shown that many of the acoustic cues to early voicing contrasts (voice bar, coda release) are already adult-like by the age of 1;6 (Shattuck-Hufnagel *et al.*, forthcoming).

The acoustic analysis of children's early speech may provide evidence for their developing phonological representations in other domains as well. Recall that Song and Demuth (2008) found compensatory lengthening on CVC targets when the word-final consonant was omitted (CVC > CV:). In a very different domain, acoustic analysis of English showed that the one child stressed her articles at early stages of acquisition, with a decrease in interval length between the article and the preceding word over time (Demuth and McCullough 2009b). Thus, by the age of 2, this child was finally beginning to treat articles as prosodic clitics rather than as independent prosodic words. Little is known about the prosodic organization of children's early productions, and how this interacts with both prosodic constraints and planning/production issues. Using acoustic and articulatory information (including the use of ultrasound measures) may help to address these issues.

5.3 *New Longitudinal Databases and Data Analysis Tools*

One of the challenges for the field of phonological acquisition has been the lack of publicly available longitudinal phonetically transcribed data from multiple children between the ages of 1 and 2 years, even for well-studied languages like English. The availability of this type of data is particularly important due to the fact that children are actively acquiring segmental, syllabic, and prosodic word structures during this time, as well as early morphology. This lack of early spontaneous production data has been especially problematic since it is often difficult to conduct elicited production experiments with children below the age of 2, and it is extremely time consuming to collect and transcribe large samples of spontaneous speech.

Beginning with the collection of semi-longitudinal semi-cross-sectional data from 12 Dutch-speaking children (the CLPF database), there is now an increasing number of phonetically transcribed longitudinal corpora from different languages. Many contain child speech from the onset of first words until around the age of 3, as well as interactions with parents, providing some indication of the nature of the input these children receive. Some of these databases also contain

linked acoustic files, facilitating verification of the phonemic transcription, and allowing for the acoustic analysis of both child and adult speech. Fortunately, many of these databases are now publicly available on CHILDES (MacWhinney 2000). A summary of some of these resources is presented in (2) (see http://childes.psy.cmu.edu/ for the latest updates and documentation/manuals).

(2) Early longitudinal data containing IPA transcription

Language	*Children*	*Ages*	*Hours*	*Database*
Catalan	4	1–4	112	Serra-Solé Catalan Corpus
Dutch	12	1–3	132	CLPF Corpus
English (American)	6	1–3	365	Providence Corpus
English (British)	2	1–3	30	Cruttenden Corpus
English (British)	1	2–3		Smith 1973 Corpus
English (Canadian)	2	2–4	30	Montréal English Corpus
French (France)	4	1–3	185	Lyon Corpus
French (Canadian)	2	2–4	30	Québec French Corpus
Japanese	3	1–2	75	Ota Corpus
Portuguese (European)	8	1–4	140	Freitas Corpus
Spanish (Spain)	1	1–3	40	Llinàs-Ojea Corpus
Spanish (Spain)	1	1–3	50	López Ornat Corpus

The collection and transcription of child speech corpora is an extremely time-consuming and labor-intensive task. The increasing availability of this type of data in the public domain will facilitate the investigation of phonological development for years to come. However, the number of children included in these studies is still very limited. Thus, there will be an ongoing need for other data sets in the future, designed to address specific phonological research questions. Although data collection and archiving, as well as the availability of Unicode IPA fonts, make this process easier than ever before, automatic transcription of the acoustic signal (forced alignment) has yet to be perfected, even for adult speech. However, in addition to scripted programs, the tool PHON is now available for conducting phonological analysis (Rose *et al.* 2006), and is freely downloadable from the CHILDES website. Acoustic analysis can also be carried out using downloadable Praat tools (Boersma and Weenink 2005). All of these developments should facilitate the collection and analysis of additional phonological acquisition corpora in the years to come, providing critical information about the initial stages of phonological acquisition, and the implications for the emergence of grammatical morphemes.

5.4 *New Experimental Methods*

Spontaneous production corpora can provide a wealth of information about children's acquisition of phonological units that occur frequently in spontaneous

speech (e.g. segments, syllable structures, and word truncation). However, it is less informative regarding the acquisition of lower frequency phonological phenomena (e.g. certain grammatical morphemes, cluster types, or word shapes). Elicited production experiments are therefore helpful in providing sufficient tokens of the phonological issue being investigated while also controlling for issues of segmental and prosodic context (though they typically cannot be used below the age of 1;6). Many of the studies discussed above have employed various elicited production techniques with both familiar and novel words, examining the constraints on children's early language productions using perceptual and acoustic measures. New methods using ultrasound technology are also beginning to examine the nature of children's articulatory gestures between the ages of 1–4 (Gick 2007; Ménard, Loevenbruck, and Savariaux 2006). These acoustic and articulatory studies hold much promise for better understanding the nature of children's early phonological representations, and how these develop over time.

Infant-speech perception studies can also be used to investigate learners' early sensitivities to different types of phonological structure (see Morgan and Demuth 1996 and Jusczyk 1997 for reviews). It has long been known that infants can discriminate between native and non-native segmental contrasts by 11 months (e.g. Best, McRoberts, and Sithole 1988). Recent studies show that, by 19 months, infants have highly sophisticated featural representations that are sensitive to changes in voicing, manner, and place, at least for consonants at the beginnings of familiar words (White and Morgan 2008). Furthermore, a series of studies has recently shown that, for fourteen-month-olds, recognizing new words presents more of a challenge than recognizing known words (cf. Swingley 2007 for a review). Infants also show a very early preference for listening to the prosodic structure of their native language (Nazzi and Ramus 2003), and can pick out high-frequency familiar words (such as their name) from the speech signal by 6 months (Bortfeld *et al.* 2005). However, identifying disyllabic words with iambic structure in the speech stream presents a challenge until 16 months for infants learning both English and French (Nazzi *et al.* 2006). Taken together, these findings suggest that the phonological representation of familiar words is more robust than that of words that have only recently been encountered.

As the infant studies have become more sophisticated, moving from segments to words and morphemes, the fields of production and perception are beginning to overlap (Gerken and McIntosh 1993). Some of the first studies that test both perception and production in the same children in a referential task indicate that the connection between the two might be tighter than often assumed (Sundara *et al.* 2011). This is obviously an area for further research.

5.5 *Modeling Phonological Learnability*

Along with developments in OT came a renewed interest in addressing phonological learnability issues. Some of the first research to examine this issue took the perspective of constraint reranking as a function of accumulating constraint violations (Prince and Tesar 2004; Tesar and Smolensky 2000). Other research has

explored constraint reranking in terms of frequency-induced changes in constraint weights (Albright and Hayes 2006; Boersma and Hayes 2001; Boersma and Levelt 2000). Recent models have begun to explore more probabilistic approaches to learning constraint rankings (Goldwater and Johnson 2003), phonological categories (Goldsmith and Xanthos 2009), and syllable structures (Goldwater and Johnson 2005; Hayes and Wilson 2008). As in any modeling enterprise, the question is always what type of input the machine learner receives, and how representative this is of what human learners have to work with. This issue is especially relevant for modeling the acquisition of phonology, where the machine learner is often trained on segmented words, whereas the human learner must identify words from the unsegmented acoustic representation of the speech stream. Thus, there is still much to be done in terms of developing more psycholinguistically plausible models of phonological learning.

6 Conclusion

The field of phonological acquisition has grown significantly since the development Optimality Theory. This has provided a much-needed framework for exploring the nature of the constraints on early phonological grammars, and how these change over time. It has also coincided with an increase in the number of phonetically transcribed corpora from different languages containing longitudinal spontaneous production data from children between the ages of 1–3. This has begun to allow the field to make and test predictions about the factors that influence the acquisition of phonology, and how this develops over time. In conjunction with new experimental methods investigating perceptual, acoustic, and gestural aspects of phonological development, and more sophisticated means of modeling the learning process, the future of the field is open-ended. New findings should help provide a clearer picture of how phonological systems are acquired, with implications for better understanding the nature of phonological disorders, the evolution of language, phonological change, and possibly phonological theory itself.