Motor Influences on Grammar in an Emergentist Model of Phonology

Tara Mc Allister Byun1* and Anne-Michelle Tessier2

1Department of Communicative Sciences and Disorders, Steinhardt School of Culture, Education and Human Development, New York University, New York, NY USA
2Department of Linguistics, Simon Fraser University, Burnaby, B.C., Canada USA

Any account aiming to provide a comprehensive picture of children’s acquisition of speech must consider both the development of the phonological grammar and the maturation of the structures and motor skills used to implement the sounds of a language. Much previous literature has been marked by a tendency to draw sharp demarcations between motor and grammatical influences, or to assert that all of child speech can be reduced to one or the other. This paper argues that it is neither necessary nor desirable to segregate speech–motor development from grammatical development when modeling speech acquisition, because they are fundamentally intertwined. The paper focuses on bringing together two literatures that have evolved largely independently. The first explores how speech–motor patterns practiced during babbling come to be disproportionately represented in the lexicon in children’s earliest stages of meaningful speech. The second posits that abstract elements of phonology – segments, features, and constraints – can be understood to emerge from generalizations over stored memory traces at a more holistic level. We argue that an emergentist model of phonological learning can be enhanced by incorporating the insight that memory traces of strings that have been heard and produced are encoded more robustly than strings that have only been heard.

1. Introduction

As anyone who has spent time in the company of a child can affirm, children’s speech output differs from that of adults in both systematic and sporadic ways. The nature and origin of these differences has been the subject of extensive debate, particularly around the relative importance of competence versus performance in shaping child speech. Some accounts frame children’s speech patterns as permutations of the same grammatical building blocks that make up adult phonologies, such as constraints or rules (Smith 1973; Goad 1997; Morissette, Dinnsen, and Gierut 2003; Gnanadesikan 2004; Pater and Barlow 2003; Dinnsen 2008 inter alia). An alternative approach emphasizes extragrammatical performance pressures – immaturity in the development of speech structures and motor control – as the major source of deviation between child and adult speech (e.g., Hale and Reiss 1998, 2008). This paper synthesizes results from phonology, behavioral psychology, and neuroscience to make the case that it is neither necessary nor desirable to draw a line between performance and grammar in the acquisition of spoken language production: rather, they are expected to be fundamentally intertwined.

The paper begins by reviewing an accumulation of evidence that grammatical and performance-based patterns are inextricably interrelated in child speech development. In the interest of brevity, this review will focus on production-oriented performance pressures – that is, factors pertaining to anatomical and motor control differences. Child speech patterns are shaped not only by widely shared biases in the development of motor control (e.g., MacNeilage and Davis 1990) but also by each individual speaker’s previous history of speech production, including preferences in prelinguistic babbling (e.g., Stoel-Gammon and Cooper 1984). The latter
line of evidence links to a recent literature suggesting that speech strings that have been articulated are represented more robustly in memory than strings that have been heard but not produced (e.g., Keren-Portnoy, Vihman, DePaolis, Whitaker, and Williams 2010). This asymmetry, which finds a plausible neural basis in the dual-stream model of speech perception (Hickok and Poeppel 2004, 2007), can account for the tendency for a child’s early lexicon to be dominated by strings that he or she can approximate in production.

We then connect these findings to a growing consensus that phonological knowledge involves both phonetically detailed episodic traces and abstract categories such as segments and features, with the latter emerging out of generalizations over the former (e.g., Munson, Edwards, and Beckman 2005; Pierrehumbert 2003; Werker and Curtin 2005). Adopting an emergentist model of phonology means that the production influences known to shape early lexical selection and production preferences can also become grammatically encoded, and so should surface in more classically phonological developmental patterns. As the grammar begins to extract segment-level representations from word-level traces, the frequently occurring, robustly encoded sounds that the child has produced should continue to be overrepresented relative to sounds that the child has rarely or never produced. If features then emerge as generalizations over natural classes of segments, it follows that the earliest-emerging features will also be biased in favor of the early production repertoire (Levelt and van Oostendorp 2007). Finally, if constraints on feature combinations are similarly emergent, then the earliest-emerging constraints must also necessarily make reference to the earliest-emerging features. In this way, the production biases that shape the earliest stages of lexical development can propagate up through higher levels of abstraction.

By examining the predictions that arise when the production advantage for learning is situated within an emergentist model of phonology, we link two literatures that have evolved largely separately. While previous works have implicitly or explicitly anticipated the intersection of these lines of research (e.g., Curtin and Zamuner 2014; Fikkert and Levelt 2008; Majorano, Vihman, and DePaolis 2014; Stoel-Gammon 2011; Vihman and Velleman 2000), no account has to our knowledge explicitly laid out the predictions generated when they are brought together. The overarching aim of this paper is to make a comprehensive case that incorporating individual-specific production experience into the construction of grammatical representations does not entail abandoning abstract grammar or its role in child speech. Instead, it enhances our ability to account for data that have proven problematic for purely innatist formal models, while situating phonological development within the broader context of cognitive development and motor maturation.

1.1. Phonological and Physical Influences on Child Speech

1.1.1. Target, Innovative, and Child-specific Patterns

It is uncontroversial that child speakers demonstrate emerging control of restrictions and alternations that are unambiguously classified as phonological in adult grammars. For example, Turkish-acquiring children as young as 1;3 (years;months) can correctly choose between [−a] versus [−ə] suffixes according to the language’s system of vowel harmony (Aksu-koç and Slobin 1985). Similarly, MacWhinney (1975, 1978) documented the early emergence of morphophonological alternations in Hungarian. The fact that children can apply these alternations in nonword contexts demonstrates that they reflect the workings of the grammar and are not mere surface imitations of adult pronunciations.

However, studies of developmental phonology have more often focused on systematic sound patterns in child speech that are not modeled by adult speakers in their environment, such as
patterns of stopping in which child speakers replace target fricatives with stops at the corresponding place of articulation (e.g., Smith 1973). Many such patterns are easily understood in the broader context of adult phonological typology. Although Standard American English does not feature a phonological pattern of stopping, there are numerous languages in which stops alternate with fricatives in contexts that are less tolerant of marked segments, like syllable codas in Korean (Ahn 1998). Another frequent innovation found in child speech is the truncation of longer words to a disyllabic or one-foot maximum (e.g., Adam and Bat–El 2009; Fikkert 1994; Tessier 2015); this particular size restriction is also attested in adult phonologies, as in English hypocoristics and in Lardil reduplication (e.g., McCarthy and Prince 1986).

Still more striking are child phonological patterns that are not only unattested in the language of the child’s environment, but in fact lack counterparts in all of adult phonological typology. Among the most thoroughly studied patterns in developmental phonology (see, e.g., Pater 1997, 2002; Pater and Werle 2003; Goad 2004; Becker and Tessier 2011) is a child-specific consonant harmony pattern involving assimilation with respect to major place of articulation (e.g., bug → [gag]; top → [pap]). Adult languages may also feature patterns of nonlocal consonant assimilation (e.g., Hansson 2001; Rose and Walker 2004; Shaw 1991), but these are specific to secondary place features, notably the features that specify the position of the tongue tip/blade in the articulation of coronal consonants (Rose and Walker 2004). In fact, adult grammars consistently prefer that non-adjacent consonants disimulate for major place (Coetzee and Pater 2008; Greenberg 1960; McCarthy 1986; Walter 2007). Another well-known example, positional velar fronting, is discussed in some detail in the following section. Such child-specific phonological patterns are problematic to model in frameworks that assume continuity between child and adult grammars (e.g., Macnamara 1982; Pinker 1984). That is, if we make the assumption that mature and developing phonologies reflect different permutations of the same basic elements, it is difficult to capture the full range of child speech patterns without generating incorrect predictions for the possible range of grammars in adult typology (see extensive discussion in Tessier 2015).

1.1.2. Motor Pressures in Child Speech Patterns

At the same time that children are acquiring native-language phonology, they are also undergoing substantive changes in the shape and control of the structures used in speech production. The vocal tract undergoes a major reconfiguration in the first two years of life (Crelin 1987; Vorperian et al. 2005), with more incremental but still rapid growth through childhood and adolescence. Children’s speech-motor planning capabilities also expand rapidly from infancy through the preschool years (Green, Moore, Higashikawa, and Steeve 2000; Kent 1992), with ongoing refinement into adolescence and even early adulthood (Goffman and Smith 1999; Romeo, Hazan, and Pettinato 2013). Numerous accounts have argued that areas of divergence between child and adult phonological patterns can be traced back to the distinct phonetic pressures arising from these motor and anatomical differences. We will illustrate with the well-studied example of positional velar fronting (e.g., Ingram 1974; Chiat 1983; Bills and Golston 2002; Inkelas and Rose 2007, 2003; Dinnsen 2008; McAllister Byun 2012). In velar fronting, consonants with a target velar place of articulation are realized with a coronal place of articulation; in the positional variant, this fronting applies in foot-initial contexts (i.e., word-initially or in the onset of a stressed syllable), but not elsewhere. Examples of positional velar fronting, taken from Inkelas and Rose 2007, are provided in Table 1.

Inkelas and Rose 2007 argued that the larger size and more anterior placement of the tongue in early stages of development, as noted above, provide the phonetic seeds for the process of positional velar fronting. McAllister Byun (2012) expanded on motor control factors that could contribute to the pattern. In the earliest stages of production, child speakers are reported to
go through a period in which the tongue plays a relatively passive role in articulation, borrowing its movements from the active jaw articulator (e.g., Green et al. 2000; MacNeilage and Davis 1990; but see Giulivi, Whalen, Goldstein, Nam, and Levitt 2011). Children as old as four have been found to show a persisting bias favoring speech sounds in which the tongue and jaw function as a single unit (e.g., Edwards et al. 1999), and such gestures can produce undifferentiated linguopalatal contact that extends across alveolar, palatal, and velar regions (Gibbon 1999). Perceptually, children’s undifferentiated gestures can be classified as velar, alveolar, or something in between (Gibbon 1999; Munson, Johnson & Edwards 2012), but these perceptual consequences tend to be stable within a given child (Gibbon & Wood, 2002). In this account, velar fronting would be observed in the subset of children who habitually release the undifferentiated closure in the coronal region, yielding a [t]-like percept. The positional nature of velar fronting has been argued to reflect the fact that the magnitude of gestural excursion is larger in an initial or prosodically strong context, and the motor control task of producing a discrete lingual gesture is correspondingly more difficult (McAllister Byun, 2012).

### Table 1. Examples of positional velar fronting from English-acquiring child “E” (Inkelas and Rose, 2007).

<table>
<thead>
<tr>
<th>Target</th>
<th>Target IPA</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>cup</td>
<td>[tʰʌp]</td>
<td>1:09.23</td>
</tr>
<tr>
<td>again</td>
<td>[sʰdin]</td>
<td>1:10.25</td>
</tr>
<tr>
<td>bagel</td>
<td>[bɛɡju]</td>
<td>1:09.23</td>
</tr>
<tr>
<td>back</td>
<td>[bæk]</td>
<td>1:10.02</td>
</tr>
<tr>
<td>coffeemaker</td>
<td>[tʰafɪmejka]</td>
<td>1:10.14</td>
</tr>
<tr>
<td>hexagon</td>
<td>[hɛksa, dən]</td>
<td>2:02.22</td>
</tr>
</tbody>
</table>

1.2. Child Speech Patterns are More Than “Pure Performance”

Both Inkelas and Rose (2003, 2007) and McAllister Byun (2012) emphasized that although positional velar fronting shows clear roots in anatomical and motor control pressures, these patterns should nonetheless be analyzed as part of the child speaker’s phonological grammar. It is important to note the contrast between this “phonetically grounded phonology” approach and a competing account that characterizes child speech patterns as purely the product of performance limitations, divorced explicitly from grammatical competence. From this latter perspective, Hale and Reiss (1998, 2008) have argued that child-specific phonology can be equated with “pseudo-phonological” processes in adult speech that are uncontroversially performance based, such as patterns of deaffrication and final devoicing in the speech of adults under the influence of alcohol (Hale and Reiss 1998; p. 669: citing Johnson, Pisoni, and Bernacki 1990). In response, Inkelas and Rose (2003, 2007) argued that intoxication can hardly be expected to produce the systematic conditioning of the kind found in positional velar fronting, where velar targets take on alveolar place exclusively at the left edge of a prosodic foot. They further emphasized that the distribution of positional velar fronting mirrors the distribution of an adult allophonic pattern, obstructive strengthening in prosodically strong contexts (Fougeron and Keating 1997; Keating 1995). Children with positional velar fronting are thus demonstrating phonological knowledge of a prosodically conditioned alternation, but the nature of the alternation that they produce reflects physical pressures specific to child speech.
Beyond the specific context of positional velar fronting, the assertion that child speech patterns are purely the product of performance limitations has been challenged in a substantial literature arguing that developmental patterns show all of the hallmarks of phonological grammar (e.g., Smith 1973; Fikkert 1994; Freitas 1997; Barlow 1997; Rose 2000; Goad & Rose, 2004). Natural class effects represent a further source of evidence that child speech patterns apply at an abstract, grammatical level of representation. Accounts of the child speech pattern of fricative stopping often invoke the high level of articulatory skill required to maintain a narrow constriction for fricative airflow without overshooting into a stop closure (McAllister Byun 2011). In many children, stopping affects fricatives across multiple places of articulation, such as labial and coronal. Since the structures and muscles used to produce these different fricatives are largely distinct, in a purely performance-based account, the elimination of stopping subtypes are independent events that are not necessarily expected to co-occur in time. However, the literature provides examples of children who, when eliminating the pattern of fricative stopping, apply the change across all places of articulation in a virtually simultaneous fashion (e.g., Rose 2014). Such cases suggest that the stop–fricative distinction is being represented at a more abstract level, transcending the individual structures and gestures involved.

A final piece of evidence supporting the grammatical status of child speech patterns comes from the robustly documented phenomenon of U-shaped trajectories in phonological development (e.g., Leopold 1939; Leopold 1947; Ferguson and Farwell 1975; MacWhinney 1978; Bowerman 1982; Fikkert 1994; Bernhardt and Stemberger 1998; Inkelas and Rose 2007, 2003; Becker and Tessier 2011; Rose and Brittain 2011; McAllister Byun, 2012). In U-shaped learning, a child is observed to produce a speech target with relatively high accuracy early in development, then shift to a pattern of incorrect production that may be sustained for some time before accuracy again increases. Because the children in question have previously demonstrated themselves capable of approximating the adult target, U-shaped learning curves pose a challenge for any account that casts child speech errors in terms of a physical inability to attain a target.

2. How does production experience shape phonology?

2.1. Empirical Evidence for a Production-Phonology Link

Some child-specific speech patterns, like the above-described example of positional velar fronting, appear to reflect the influence of anatomical and motor constraints that are broadly shared across child speakers. In some theories, these shared physical pressures are encoded in an innate hierarchy of markedness (Jakobson, 1941/68) and thus could be considered purely grammatical. However, there is evidence that a child’s individual production experience can have a systematic impact on subsequent phonological development, as manifested in the lexical selection and template effects reviewed in this section.

Lexical selection and lexical avoidance are observed early in development when children appear to restrict which words they attempt based on the phonological characteristics of those items (e.g., Ferguson and Farwell 1975; Menn 1983; Schwartz and Leonard 1982; Stoel-Gammon 2011; Vihman 2014). In lexical selection, children disproportionately attempt to produce words whose sounds or syllable shapes fall within their current production inventory; in lexical avoidance, children refrain from attempting lexical targets containing sounds or sequences outside that current inventory.

Given our present focus on links between articulatory experience and phonological development, it is noteworthy that lexical selection has been shown to reflect speakers’ previous history of pre-linguistic babbling (Bernhardt and Stoel-Gammon 1996; Stoel-Gammon and Cooper 1984; Vihman, Macken, Miller, Simmons & Miller, 1985). Although some recurring patterns
in babbling have been attributed to universal biomechanical pressures (e.g., MacNeilage and Davis 1990), infants also show individual preferences for particular syllable shapes, places or manners of articulation. These preferences are also at least partly independent of the frequency of occurrence of different segments in the maternal input (DePaolis, Vihman, and Keren-Portnoy 2011; DePaolis, Vihman, and Nakai 2013). Longitudinal research has established that children who favor a particular sound or syllable in their babbling tend to carry this same preference into their early word productions (e.g., Bernhardt and Stoel-Gammon, 1996; Stoel-Gammon and Cooper 1984; Vihman et al., 1985). For example, Stoel-Gammon & Cooper (1984) described a child whose pre-linguistic babbling reflected a strong preference for CVC syllables with velar place (e.g., [gak]). An analysis of his first 50 words revealed that 22% featured a final velar stop, versus only 4-8% final velars in the lexical repertoires of other children in the paper. Although lexical selection and avoidance are most strongly associated with the onset stages of speech, they can also be observed in some children with larger vocabularies (Stoel-Gammon, 1998, 2011; see also Donahue 1986.)

As children’s vocabularies grow and they attempt more forms that diverge from their preferred shapes for production, learners also actively alter adult target forms to achieve a closer match with their preferred production patterns. In some cases, these preferred patterns are best described as whole-word targets or templates; there is no readily identifiable segment-by-segment mapping between the adult form and the child’s output (Waterson 1971; Priestly 1977; Leonard and McGregor 1991; Macken 1996; Vihman and Velleman 2000; Vihman and Croft 2007; Menn, Schmidt, and Nicholas 2009, 2013; Wauquier and Yamaguchi 2013; Vihman 2014). A well-documented English example comes from Priestly (1977), reporting on one child, C, who imposed a [CVjaC] template on many bisyllabic word targets for several weeks beginning at 1;11. The diary paper further demonstrated that in the first week that this template became prevalent, lexical targets with medial [j] were disproportionately frequent in C’s output; in the dialect C was acquiring, this included words such as lion, wheel, nail, and fire. This exemplifies a trajectory described by Vihman and Croft (2007): after stabilizing a speech-motor routine through repeated production of one or more adult targets that broadly conform to a particular shape, the child extends this template to encompass a wider range of target words. In a child who makes extensive use of templates, some outputs may seem to lack a systematic relationship to the adult target, yet show a transparent link to other words in the child’s output (Macken 1979; Vihman and Croft 2007; Vihman and Velleman 2000).³

2.2. Mechanisms Supporting a Production-phonology Link

Although the influence of early production experience on later output patterns is now amply documented, the mechanism by which production experience facilitates later learning is less thoroughly understood. The leading view holds that lexical acquisition is facilitated when children possess a stable articulatory routine, also termed a “vocal motor scheme” (DePaolis et al. 2011; DePaolis et al. 2013; McCune and Vihman 2001; Stoel-Gammon 2011), in connection with a particular sound sequence. Many components can be identified in the task of learning a word for production purposes: encoding its auditory-acoustic form, mapping this form to a stored meaning, and also accessing the speech-motor routine that best approximates the form’s auditory target. If the word corresponds with a well-practiced speech string, this last component will be facilitated, reducing the overall processing load and increasing the likelihood of successful word-learning (DePaolis et al. 2011; McCune and Vihman 2001). Nonce word learning experiments in children and adults have supported the hypothesis that experience producing a sound sequence makes it easier to acquire a word-meaning mapping incorporating that sequence (Kan,
Evidence that previously produced strings are represented more robustly in memory than strings that have only been heard goes beyond the specific task of acquiring new lexical meanings. For example, Keren-Portnoy et al. (2010) found that 26-month-olds were significantly more accurate at repeating nonwords constructed from consonants that they produced frequently, compared to consonants that were also developmentally appropriate but were used less frequently by the child in question. Ettlinger, Lanter and Van Pay (2014) found that three- to five-year-old children were more accurate in recalling the plurality of a seen object when the plural was easily pronounceable (e.g. shoe versus shoes) than when the plural contained a challenging consonant sequence (e.g. sock versus socks). This effect was obtained even though overt speech was not elicited during either the exposure or the recall phase.

Production-related asymmetries in the perceptual encoding and retention of speech have a plausible neurobiological basis in the dual-stream model of speech perception (Hickok and Poeppel, 2000), which posits that speech inputs are processed along two anatomically and functionally distinct neural pathways. The *ventral* stream links auditory sensory input to associated semantic representations and is used in everyday tasks of understanding speech. The *dorsal* stream, which extends from the temporal-parietal border to the posterior frontal lobe, makes it possible to map between an auditory representation of a sound sequence and the motor regions associated with production of that sequence. The dorsal stream can thus be presumed to play a vital role in speech acquisition, which at its core is a task of learning what movements of the vocal tract will produce the closest match for an adult auditory target (Doupe and Kuhl 1999; Guenther, Hamson, and Johnson 1998). Activity in the dorsal stream is also observed in adults during tasks involving the sub-vocal rehearsal component of verbal working memory (Awh et al. 1996; Buchsbaum, Hickok, and Humphries 2001; Ylinen et al. 2014), as well as during speech perception in noise and other effortful processing conditions (Adank and Devlin, 2010; Utanski, Cavinetes, and Liss 2015). Of importance for the present discussion is the finding that dorsal stream activity is also crucial for tasks that require detailed, sublexical levels of speech processing, such as discriminating syllables that differ by a single phoneme (e.g., Baker, Blumstein, and Goodglass 1981; Hickok and Poeppel 2004; Miceli, Gainotti, Caltagirone, and Masullo 1980).4

With this groundwork in place, we return to the idea of vocal motor schemes as “relatively robust sensorimotor patterns...available for use in phonological memory” (Majorano et al., 2014: 183). One line of longitudinal research using the headturn preference paradigm (Nelson et al.,1995) has accumulated some initial evidence for an association between production experience and perceptual processing in prelexical infants (DePaolis et al. 2011; DePaolis et al. 2013; Majorano et al. 2014). A general finding that has been consistently supported across these studies is that greater experience producing a consonant in babble is associated with a decrease in looking time to speech containing that consonant, relative to speech featuring unpracticed consonants. Under the assumption that infants look longer to streams of speech that are more challenging for them to process, this result suggests that perceptual processing is less effortful for inputs that have previously been produced (DePaolis et al. 2013). Recent neuroimaging work with infants at a similar age provides a striking parallel. Kuhl, Ramírez, Bosseler, Lin, and Imada (2014) found that seven-month-olds show high levels of activity in both superior temporal (auditory) and inferior frontal (motor) regions of the brain while listening to any speech signal. Adults, by contrast, show an asymmetry whereby non-native speech stimuli trigger a higher level of activity in motor regions of the brain than native speech sounds. By twelve months, after infants have accumulated extensive babbling experience, their brain responses show this same asymmetry.

These strands of research jointly create a coherent picture of a mechanism underlying lexical selection and other influences of production experience on phonological development. When
children hear speech strings that they have little or no experience producing, it can be hypoth-
esized that the dorsal pathway is engaged in the search for the motor command that will most
closely map to that auditory string. This is reflected in the inferior frontal activation observed
by Kuhl et al. (2014), as well as the behavioral evidence of effortful processing in Majorano
et al. (2014). By contrast, when children hear speech strings that they have produced in the past,
they can more efficiently retrieve the associated motor representation, and the dorsal pathway
can be devoted to its other functions: refreshing memory traces via subvocal rehearsal and
encoding them with a higher level of sublexical detail. It is reasonable to posit that the robust
traces that result are better able to support word learning and subsequent stages of processing,
discussed below.

2.3. From Vocal Motor Schemes to Phonological Grammar

The connections drawn in this section – among production experience, the early lexicon, and
the robustness of speech strings in memory – have been based on evidence from the very early
stages of speech development, when children’s production patterns are more likely to reflect id-
iosyncratic templates than systematic segmental or prosodic patterns of the type seen in adult
phonology. As discussed in Section 1, the influence of production on child phonology is not
exclusive to these early stages. Patterns like positional velar fronting can become a productive
part of a child’s phonology, applying to new words and nonwords as well as existing lexical
items, and they may continue in children three to four years of age or even older (e.g.,
McAllister Byun, 2012). The finding that production experience can enhance memory
encoding does not trivially extend from the word or syllable level to more abstract patterns.
Fortunately, it is possible to shed some light on this issue by consulting an entire literature ex-
ploring the subject of emergent structure in phonology. Research in this tradition investigates
how lexical learning and sublexical phonological development can be understood to be mutually
interdependent, with more abstract units emerging from detailed episodic traces. In the discussion
below, we aim to connect the evidence for individual-level production influences with this lit-
erature on emergent phonological grammar, laying the framework for an account of motor in-
fluences on child phonology that extends beyond the earliest stages of lexical development.

3. Emergentism in phonological acquisition

3.1. Why Emergentism?

The literature on speech processing in adults has converged on a consensus that spoken informa-
tion must be processed and represented at multiple levels of abstraction. There is abundant evi-
dence that speakers are sensitive to episodic detail, such as speaker identity and affect, and
equally robust evidence that speakers also process information at the level of abstract categories
(Pierrehumbert 2016). Studies of speech perception in infants and children have demonstrated
that they too are sensitive to both episodic (e.g., Nielsen 2014; Singh, Morgan, and White
2004) and categorical (e.g., Eimas and Miller 1980; Eimas, Siqueland, Jusczyk, and Vigorito
1971; Jusczyk 1977) properties of speech. As a consequence, recent models of phonological learn-
ing tend to include both a detailed episodic level of representation and one or more higher levels,
with each layer “abstracted progressively further away from the parametric phonetic encodings”
(Munson, Edwards, and Beckman, 2005: 299; in another vein, see Jarosz, Calamaro & Zentz,
under review; Kiparsky and Menn, 1977). The classic finding that the number of abstract phono-
logical categories posited by an infant is sensitive to the distributional properties of the input
(Maye, Werker, and Gerken 2002) provides evidence that the abstract levels of representation
emerge out of the phonetic detail – as opposed to, say, the episodic detail being mapped onto a predetermined set of innate categories.

This process of accumulating a large number of observations and inferring a category from their patterns of clustering and separation is the most fundamental type of emergent abstraction (e.g., Pierrehumbert 2003). One complication stems from the fact that, while distributional patterns may be easily observed when inputs are discrete points in a multidimensional space, children rarely encounter speech sounds in isolation; their input consists primarily of strings of syllables or words. This implies that emergent phonology must involve extracting smaller units out of larger speech chunks, as suggested by the large and long-established literature exploring “the primacy of lexical items in phonological development” (Ferguson and Farwell 1975: 437; see other early discussion in Francescato 1968; Menn 1971; Waterson 1971; Ferguson, Peizer, and Weeks 1973). The sections that follow will expand on the evidence for an emergentist model of phonological grammar, beginning with the emergence of segments from word-level representations and proceeding to the emergence of features and constraints.5

3.2. Emergence of Segments

Although the present paper focuses on children’s speech production abilities, it is of course strongly connected to previous proposals emphasizing development in the perceptual domain. One prominent and comprehensive account is PRIMIR (Processing Rich Information from Multidimensional Interactive Representations: Werker and Curtin 2005; Curtin, Byers-Heinlein, and Werker 2011), a psycholinguistic model of perception which specifically posits the emergence of segments from a holistic word-level tier. In the PRIMIR framework, language learners detect and organize information in three dimensions or spaces; different dimensions can take on greater salience, depending on the task or the developmental stage of the individual (Werker and Curtin 2005). Phonetically detailed exemplars are stored in General Perceptual and Word Form spaces, which gradually take on internal structure as tokens cluster and overlap in multidimensional acoustic space. Eventually, the child’s identification of meaningful regularities in the sound system (i.e., differences in phonetic form that are associated with lexical contrasts) allows him/her to project a finer-grained Phoneme space (see Pierrehumbert 2002 for a similar account). Although no articulatory information is considered in the PRIMIR model, there is increasing consensus that a comprehensive psycholinguistic model will need to encompass not only lexical and phonetic information but also more distal associations such as the discourse context of the utterance or the social status of the speaker (Fink & Goldrick, 2015; Pierrehumbert 2016). In the context of such massively multidimensional representations, it seems simple to suggest that speakers may additionally encode information about the motor actions associated with known speech sequences (for a related concept, see McMurray and Farris-Trimble’s (2012) account of “emergent information-level coupling between perception and production”).

3.2.1. Evidence for Holistic Early Representations

Various properties of children’s early outputs support the contention that lexical representations start out holistic and lacking in segmental detail. Ferguson and Farwell (1975) emphasized the variable character of early word attempts, highlighting a fifteen-month-old child whose efforts within a single recording session to produce the word “pen” ranged from [dɛdn] and [hɪn] to [mãə] and [mbô]. This variability suggests that the speaker is aware of the general acoustic properties of the lexical target, but this information is not yet linked to specific structural subunits such as segments (see also the sound-level variability discussed in Vihman and Croft 2007).
The strong tendency toward place assimilation in children’s earliest productions, which affects both consonants and vowels, is also consistent with the hypothesis of holistic early representations. Fudge (1969) described a 1;4-year-old English-learning child whose realization of major obstruent place was harmonized with following vowel: labial and velar targets were produced as coronals before front vowels, while target coronals became labial before back rounded vowels and velar before back unrounded vowels; corpus data along similar lines from Fikkert and Levelt (2008) is discussed below. A final piece of evidence comes from the template or “whole-word processes” discussed in Section 2.1. While older children’s phonological processes can typically be described with reference to substitution, deletion, or insertion of phonemes or syllables, templatic outputs appear to be targeting the gestalt of the word rather than sublexical elements (Stoel-Gammon 2011; Vihman and Croft 2007; Waterson 1971).

It is important to acknowledge that children’s early representations cannot be completely devoid of sub-lexical detail, since infants at around one year of age show sensitivity to single-segment mispronunciations of familiar words such as “tog” for dog (Swingley 2005; Swingley and Aslin 2002). However, the availability of this fine-grained information is heavily dependent on the task and the familiarity of the lexical item; this supports a PRIMIR-type view in which infants’ knowledge of segments is represented at multiple levels and may be accessible in differing degrees across different contexts (Curtin and Zamuner 2014).

3.2.2. Evidence for the Emergence of Segmental Representations

The preceding section reviewed evidence that in the earliest stages of meaningful speech production, between roughly one and two years of age, children’s lexical representations are less detailed than adults’. A body of research using word and nonword repetition tasks has contributed evidence that representations continue to become increasingly segmental as children acquire larger vocabularies in subsequent years of development (Edwards, Beckman, and Munson 2004; Zamuner, Gerken, & Hammond, 2005; Munson, Edwards, & Beckman, 2005). Young children’s attempts to repeat a nonword are influenced not only by the component segments but also by the frequencies with which strings of segments co-occur in the lexicon. That is, they are significantly more accurate in repeating a nonword when it contains a high-probability sequence like /ft/ (which occurs in many words, such as after and softly) as compared to a low-probability sequence like /fk/ (Edwards et al. 2004). As a child’s lexicon expands, however, nonword repetition accuracy is less strongly influenced by these transitional probabilities. This suggests that the children with large vocabularies have built more independent segmental representations — i.e., that the unit /f/ has been abstracted away from common sequences such as /ft/ and /fr/, so that co-occurrence with those following segments is less crucial for accurate production. Furthermore, a series of analyses indicate that sensitivity to phonotactic probability in nonword repetition is most directly predicted by vocabulary size, not age (Edwards et al. 2004), and not phonetic perceptual acuity or articulatory accuracy (Munson et al. 2005).

The finding that sensitivity to phonotactic probability changes with vocabulary size not only provides evidence that segmental structure emerges from more holistic representations but also points to a plausible mechanism for this transition, in line with Metsala and Walley’s (1998) Lexical Restructuring Hypothesis. Phonotactic probability and neighborhood density are correlated, such that words with high overall phonotactic probability tend to occupy dense neighborhoods, while words with low phonotactic probability tend to have fewer similar neighbors (Vitevitch, Luce, Pisoni, and Auer 1999). In high-density neighborhoods, there is pressure to replace more holistic representations with more detailed ones in order to preserve the ability to differentiate among similar targets (Metsala, 1997). The proposed role of neighborhood density in the emergence of segmental detail is also supported by evidence that words in
dense neighborhoods are produced with greater accuracy and lower variability than words in sparse neighborhoods (Stoel-Gammon and Sosa 2012, but cf. Ota and Green 2012, for evidence that production accuracy may be lower for onset clusters in dense neighborhoods).

3.3. Emergence of Features

Different strands of research in phonological acquisition have taken widely differing perspectives on the question of how and when featural abstraction is introduced into the grammar. The pioneering work of Jakobson (1941/68) suggested that children acquire their segmental inventory through successive emergence of phonological features and that this order of emergence is universal, commencing with the least marked and terminating with the most marked features.

By contrast, beginning at least with Smith (1973), most work in developmental phonology has made the implicit or explicit assumption that the child learner starts out with access to a complete, universal feature set. When a child fails to realize a contrast attested in the adult grammar, such as stop versus fricative, this is interpreted as a neutralizing process, rather than the absence of the feature to express that contrast (see discussion in Dunbar and Idsardi 2012). More recently, Levelt and van Oostendorp (2007) have proposed that initial segmental acquisition is followed by a gradual appearance of features, in a more flexible order. This “Neo-Jakobsonian” account, which is taken up in more detail in Section 4, is in line with a broader shift away from featural innateness and toward a model in which features emerge from observed data (Clements and Ridouane 2011; Dresher 2009; Mielke 2008; Rose 2009).

The view that features are abstracted from observed data has been pursued in the phonological literature perhaps most comprehensively by Mielke (2008). In Mielke’s Emergent Feature Theory, learners share certain innate biases, including a preference to organize information into hierarchically organized, increasingly abstract categories. When these learners are exposed to a wide variety of phonetic information in their native language, features are induced on the basis of distributional regularities in both acoustic-perceptual and articulatory dimensions of phonetic space. Lin and Mielke (2008) tested the feasibility of emergent features in a computational model, using hierarchical mixture models to identify clusters in a corpus of acoustic data linked to segment-level IPA transcription, as well as a similarly segmented corpus of articulatory (ultrasound) data. They found that hierarchically organized classes with strong parallels to traditional natural classes emerged from this clustering process: manner features tended to emerge from the acoustic data, while place features emerged from the articulatory data. An interesting parallel is provided by recent research measuring electrocorticographic signals in the superior temporal gyrus during connected speech processing (Mesgarani, Cheung, Johnson, and Chang 2014). When patterns of neural activity were analyzed in conjunction with a segmented record of the speech signal, unsupervised clustering yielded a set of manner-based classes similar to those that derived in Lin and Mielke’s (2008) analysis of acoustic-phonetic inputs.

How might we infer from behavioral data whether and when children are representing their phonological patterns in featural terms? One source of evidence comes from the kind of U-shaped error pattern which first affects one segment or sequence in the target language, and over time spreads to featurally related ones. For example, Menn (1971) reports that nasal harmony in child speaker Daniel initially affected only labial-initial clusters like broom [mʊm], but eventually targeted a much broader set of initial onsets including jump [mnʌmp] and stone [nʌmp] before it subsequently receded. Evidence for featural representations is strengthened to the extent that a child’s phonological process targets more or less all the items within a natural class, rather than individual words in piecemeal fashion (for evidence to this effect, see references in Tessier 2014; Levelt and van Oostendorp 2007; although cf. Vihman and Croft 2007). At the
point where a child’s phonological system has stabilized and relies less on individualistic templates, multiple examples demonstrate that children’s non-target productions can apply to natural classes rather than just individual segments; see, for example, the systematic treatment of [labial] in cluster reduction, fusion, and overwriting patterns documented by Gnanadesikan (2004), as well as across-the-board elimination of stopping across places of articulation (Rose 2014), discussed above.

3.4. Emergence of Constraints

Just as most models of phonological development make the assumption that even very young speakers have access to the full feature set of their language, most work in the widely used framework of Optimality Theory (OT; Prince and Smolensky, 2004) has assumed a universal, innate set of phonological constraints. However, an innate constraint set is not intrinsically associated with the evaluation metric of an OT grammar—in a well-cited early example, Hayes (1999) makes the case for an emergent constraint set, built from a learner’s direct experience of phonetic factors including articulatory difficulty and perceptual confusability. Recent literature has shown increased attention to the question of how learners could construct and rank or weight constraints to approximate the grammar of their language community (e.g., Adriaans and Kager 2010; Hayes and Wilson 2008; Heinz 2007; Kager and Pater 2012). Much of this literature, including Hayes and Wilson’s (2008) influential model, represents a proof-of-concept level of demonstration. That is, the aim is to show that the model, given a training corpus and certain building blocks and/or biases, can arrive at the constraints necessary to describe the adult phonology. These papers do not aim to model the specific learning trajectories of particular children, or indeed, of child learners as a whole.

Nevertheless, some work has evaluated the idea of an emergent constraint set against actual data from children’s early error patterns. A notable example is offered by Levelt and van Oostendorp (2007) whose “Neo-Jakobsonian” model, discussed in the preceding section, assumes that features and constraints emerge in parallel. Once a feature becomes active in the child’s phonology, it can be used to build featural co-occurrence constraints, which combine pairs of features in two schematic ways: either banning combinations of privative features ("*[F, G]") or stating implications between them ("if F then G"). These featural constraints can be further parametrized to prosodic contexts, restricting their scope to either onset or coda position or leaving them unrestricted. With this set of tools, Levelt and van Oostendorp (2007) report that their model can capture the developmental trajectories in a corpus of six children acquiring the segmental inventory of Dutch. For the most part, their analysis uses constraints that are adopted by all six children at the earliest stages (e.g., *[nasal, velar] and “if nasal then labial”; p. 167), albeit with different contextual restrictions. The differences between children seem to lie in the order in which features emerge and the order in which they revoke their overly-restrictive constraints, as they must eventually do in order to acquire the target Dutch inventory.

While such featural constraints successfully capture the patterns in this corpus, Levelt and van Oostendorp’s (2007) model defers the task of explaining why children construct these specific constraints to a “general theory of markedness” (p. 167). A likely source of emergent constraints is lexical distribution: features or feature co-occurrences which are under- or over-represented in the lexicon may be grammaticalized as a more categorical ban or obligation, respectively. However, Levelt and van Oostendorp (2007) describe learning trajectories that do not directly mirror the input frequencies of Dutch, and in some cases, children posit constraints that actively overrule the adult evidence. For example, when one child in their corpus acquires the feature [nasal], they first restrict it with the FCC “if nasal then labial,” even though the Dutch phonemic inventory includes [n] and [ŋ] as well as [m].
This question of where children’s “innovative” constraints come from is taken up more directly in Fikkert and Levelt’s (2008) analysis of major place harmony and other child-specific place restrictions in the same corpus of child Dutch. Observing patterns of substitution, harmony, and metathesis that cause labial place to be aligned with the left edge of the word (Jaeger 1997; Menn 1983), Fikkert & Levelt (2008) posit that children in their corpus – and presumably Dutch children more generally – construct a constraint requiring word-initial segments to bear the feature [labial]. They suggest that this constraint represents a generalization over the child’s own lexicon, rather than the entirety of the adult input. While labial-initial words are frequent in infant-directed Dutch, they are disproportionately more frequent in statistics of the child lexicon (i.e., in the words attempted by children in the corpus, as well as in a list of typical six-year-old vocabulary). Moreover, a preference for labial-initial syllables is attested in child speech in various languages (e.g., MacNeilage and Davis 2000) and appears to have a biomechanically preferred status in general (Tsuji, Gomez, Medina, Nazzi, and Mazuka 2012). For additional discussion of constraint induction reflecting child-specific phonetic pressures, see Inkelas and Rose (2007) and Becker and Tessier (2011).

4. Motor influences in an emergent phonology

Two literatures were reviewed in the preceding sections. Section 2 examined how speech-motor patterns practiced during early vocal play come to be disproportionately represented in children’s first stages of meaningful speech. Section 3 discussed how some of the abstract elements of phonology – segments, features, and constraints – can be understood to emerge from generalizations over stored episodic traces. This section aims to unite insights from both literatures in a reassessment of data drawn from the Dutch child corpus described in Fikkert and Levelt (2008). Because formal modeling is outside of the scope of this paper, we narrate the ways in which motor production experience can be seen to influence successive levels of emergent abstraction in these examples.

4.1. Modeling Early Holistic Productions

As discussed above, Fikkert and Levelt (2008) report that all children in their sample start out in a “one-feature, one-word” stage. In this phase, output syllables can be characterized by a single place of articulation, generally dictated by the vowel: front vowels are realized with coronal consonants, back rounded vowels with labial consonants, and back unrounded vowels with velar consonants. The low vowel /a/ is considered neutral and can thus combine with different consonant places.

These early productions can be understood to reflect the holistic nature of children’s earliest lexical representations, prior to the emergence of well-defined segmental structure. The influence of lexical selection is apparent here: the harmonic syllables documented in Fikkert & Levelt’s (2008) corpus paper (examples in 1) correspond with consonant-vowel combinations that MacNeilage and Davis (1990) identified as biomechanically preferred and thus disproportionately frequent in babbling. Thus, these syllables are likely to have been among the well-practiced motor-sensory mappings (vocal motor schemes) from the child’s prelexical babbling experience. As discussed above, we hypothesize that when infants hear words containing these well-practiced sequences, less processing effort is required to access the stored representation with its associated articulatory routine, and the word is more likely to be added to the lexicon.

(1) “One feature” words suggestive of lexical selection in the initial vocabulary of Robin (1;5.11)

a. die “this one” → [ti]; niet “not” → [nt]; zes “six” → [sɛs]
b. pop “doll” → [pɔ]; aap “monkey” → [ap]; mamma “mommy” → [mɑmɑ]
Fikkert & Levelt (2008) describe subsequent phases of development in terms of “staged segmentation,” i.e., the gradual emergence of discrete segments from early undifferentiated syllables. The first change is the emergence of distinct consonant and vowel places within a syllable. Following models such as PRIMIR, but extending to the production domain, we posit that this change occurs when children have amassed sufficient distributional data to detect that their stored word-level traces contain areas of acoustic and motoric overlap. Data from Fikkert & Levelt’s (2008) subject Eva (examples in 2) show that she paired a coronal onset with different rhymes while still respecting the biomechanical preference to combine coronal consonants with front vowels. Recognizing a recurring motor-sensory mapping in word-initial position could allow Eva to project [t] as a segment independent of the syllable- or word-level units it has occurred in to date. Given that auditory-articulatory transformation is thought to be a prerequisite for perceptual tasks involving manipulation of individual segments (Hickok and Poeppel 2007), production experience is likely to play a key role in arriving at this segment-level representation.

(2) Eva’s harmonic syllables at 1;4.12

a. trein “train” → [tæin]; teen “toe” → [ten]; staart “tail” → [tat]

4.3. Modeling the Emergence of Features and Constraints

Another child in the Fikkert & Levelt (2008) corpus, Noortje, displays a classic U-shaped trajectory that provides an informative example for our discussion of feature and constraint emergence. At an early stage, Noortje produced full consonant harmony within syllables, so that dorsal consonants could only co-occur with other dorsals (shown in 3a). Since outputs like kik were possible, we know that she had decomposed her holistic word forms into sufficiently segmental representations that consonants were represented separately from vowels. The next step in staged segmentation was the realization that consonants can have different place specifications within a syllable.

At the same time that she was accumulating evidence leading up to the generalization that consonant harmony is not obligatory, the learner was also searching for broader generalizations that relate segments to one another or to prosodic positions. In this case, multiple types of evidence may have drawn her attention to a positional asymmetry in the behavior of velar segments. As discussed in Section 1.1.2, the biomechanical constraints that promote fronting of velars in child speech are especially active in initial position. As shown in (3b), at the same time that she was emerging from consonant harmony, Noortje exhibited fronting of initial velars. It is possible that these instances of fronting initially arose as motor performance errors (Inkelas and Rose, 2007), but once produced, these forms were represented in memory with traces of greater salience and specificity than auditory-only traces. This would serve to reinforce a trend already present in the input, whereby word-initial velar onsets have a low frequency relative to coronal or labial place (Fikkert and Levelt 2008). In one possible interpretation, Noortje’s response was to posit a constraint banning velar place in initial position. Harmonic words, which were previously realized with faithful velar place (3a), now regressed in accuracy in accordance with the grammar’s new positional constraint (3c). This U-shaped trajectory provides evidence that a trend driven by articulatory pressures was subsequently encoded in a more abstract, grammatical form.
(3) Noortje’s acquisition of dorsals

a. *koek* “cookie” → [kuk] (2;3.7); *klok* “clock” → [kɔk] (2;5.23); *kikker* “frog” → [kık] (2;2.1)

b. *grote* “big” → [dɔtə] (2;9.1); *kip* “chicken” → [tip] (2;10.12)

c. *koek* “cookie” → [tuk] (2;8.17); *kikker* “frog” → [tika] (2;8.17)

5. Conclusions

This paper has approached the search for explanations in early phonology with the hypothesis that children’s grammars are constructed under the influence of performance factors. To support this view, we have brought together insights from two broad areas of existing literature: one documenting strong links between articulatory experience and the sound structure of the early lexicon, and another arguing for the gradual emergence of finer-grained representational structure out of distributions of episodic traces. We have aimed to demonstrate that incorporating child-specific and individual-specific experience into the construction of grammatical representations does not entail abandoning abstract grammar or its role in child speech. Instead, it enhances our ability to account for data that have proven problematic for existing formal models, while allowing us to position phonological development within the broader context of cognitive development and motor maturation.

Given the scope of this endeavor, many questions are necessarily left unanswered. For example, much of our discussion regarding higher levels of grammatical abstraction remains at the level of conjecture, and the search for learner biases that underlie featural discovery and constraint construction remains a wide-open area of research. In the interest of brevity, we have also stopped short of engaging with the question of how child-specific phonological patterns like positional velar fronting are eventually eliminated. U-shaped learning suggests that children do grammaticalize generalizations from their production experience, but they also recover from these overgeneralizations. Since adult phonological typology carries no trace of these child-specific biases – recall, for example, that adult grammars prefer to *dissimilate* non-adjacent sequences of consonants with the same major place of articulation – it is not sufficient to suggest that these constraints are merely demoted due to lack of support in the input. Instead, the evidence suggests that some child-specific constraints must be deactivated or fundamentally transformed.

An intuitive explanation stems from the fact that the relative difficulty of two speech production tasks can actually invert over the course of maturation. For example, while infants produce newly acquired consonants in repetitive strings before branching out to variegated sequences (Lipkind et al. 2013), experimental evidence suggests that alternating sequences are articulatorily more efficient and thus biomechanically preferred in adult speakers (Rochet-Capellan and Schwartz 2005; Walter 2007). It is possible that constraint induction is an ongoing process, and the constraint inventory changes as the underlying phonetic pressures shift. An alternative approach holds that the grammar includes at least one constraint that is directly sensitive to an index of articulatory reliability, which could drive child-specific speech patterns and also account for typologically unusual patterns that emerge in cases of acquired damage to speech–motor regions of the brain (McAllister Byun, Inkelas and Rose, 2016).

Despite the many questions that remain to be addressed, we maintain that the integration of concepts laid out in this paper represents a step toward a more coherent picture of the relationship between speech–motor maturation and phonological development. We anticipate that new insights will become available as psycholinguistic and neuroscientific studies investigating the relationship between speech production and memory encoding are further integrated into theoretical and computational models of how phonological abstraction emerges from generalizations over experiential traces.

© 2016 The Author
Language and Linguistics Compass © 2016 John Wiley & Sons Ltd
Short Biographies

Tara McAllister Byun’s research aims to understand how articulatory and perceptual factors shape speech development in children with and without phonological delay or disorder. She also maintains an active line of research in clinical linguistics, where she investigates the efficacy of interventions that use technologies (e.g., ultrasound and acoustic analysis) to provide visual biofeedback for persistent speech sound errors in children. She has published in journals including Language, Phonology, Journal of Child Language, Journal of Speech, Language, and Hearing Research, and Journal of Communication Disorders. Her research is supported by the National Institute on Deafness and Communication Disorders. McAllister Byun is an assistant professor in the department of Communicative Sciences and Disorders at New York University. She holds an AB and AM in Linguistics from Harvard University, an MS in Speech and Hearing Sciences from Boston University, and a PhD in Linguistics from the Massachusetts Institute of Technology.

Anne-Michelle Tessier’s training is in formal phonological theory, with a focus on constraint-based acquisition and learnability. Her research expertise lies in combining grammatical insights from cross-linguistic typologies with longitudinal data from developing child phonologies, as well as in the use of artificial language learning methods for testing theoretical predictions about biases in (morpho-) phonological learning. Her work has been published in venues including Phonology, Natural Language and Linguistic Theory, Language Acquisition, Bilingualism: Language and Cognition, and Lingua, and her textbook Phonological Acquisition: Child Language and Constraint-Based Grammar was published by Palgrave Macmillan in 2015. Recently, her research has begun to focus on early L2 phonological development, supported by an Insight Grant from the Social Sciences and Research Council of Canada. Anne-Michelle received a BA Honours in Linguistics from McGill University and a PhD in Linguistics from the University of Massachusetts, Amherst. From 2006–2016, she was faculty in Linguistics at the University of Alberta; she is now Adjunct Professor at Simon Fraser University.

Notes

1 Note that we use the term “child-specific phonological pattern” to refer to patterns that are specific to children, in the sense that they are attested in child phonology but lack counterparts in adult phonological typology. The same term has also been used to refer to an idiosyncratic phonological pattern that is specific to one child. These definitions are not mutually exclusive; note that the latter type of phonological pattern is a subset of the former.

2 Recall that while the present paper focuses on the production domain for the sake of brevity, a complete account of child-specific phonetic pressures would also consider perceptual differences between child and adult speakers.

3 Kehoe (2015) points out that there currently exist no large-scale studies systematically documenting the prevalence of templates and/or lexical selection effects. Until such research is conducted, we are unable to determine whether these phenomena are common (or even universal), or if they characterize only a specific subset of children.

4 The observation of activity in motor cortex during speech perception is sometimes interpreted as evidence in support of the motor theory of speech perception, which posits that perceiving speech directly activates articulatory gestures (e.g., Liberman and Mattingly 1985). However, this is by no means a necessary conclusion; other accounts interpret this frontal activity as a means to enhance auditory processing under challenging conditions (e.g., Adank and Devlin 2010), or as a reflection of coactivation of verbal working memory (e.g., Buchsbaum et al. 2001).

5 While in this paper we represent sublexical detail in terms of segments and features, our overall approach is equally compatible with a theory where the abstract units of phonology are articulatory gestures (as in Browman and Goldstein 1986 and subsequent work).
6 Levelt and van Oostendorp (2007) remain agnostic as to whether features become active because they are constructed anew by each child from experience, as assumed here, or whether their “emergence” is the result of innate constraints which are re-ranked to reveal their influence. The “revelation” of an innate feature through constraint re-ranking would simply mean moving from *F >> Ident-F to the reverse, Ident-F >> *F; a feature has emerged when it is tolerated in optimal candidates.

7 It should be noted that the “observed sound patterns” referenced in Mielke (2008) are derived not only from phonetic distributional patterns but also from “phonologically active classes,” which are found primarily in alternations. Since using an alternation requires sufficient knowledge of paradigms to recognize the allomorphs whose common form is undergoing a process, much of this data must be inaccessible to very young learners.

8 We follow Werker and Curtin (2005) in positing that the earliest level of abstraction that children represent is better described as a position-specific allophone; awareness of the more abstract phonemic representation emerges later.

Works cited


Jarosz, Gaja, Shira Calamaro and Jason Zentz. (under review) Input frequency and the acquisition of syllable structure in Polish.


