Book Section:
Vihman, Marilyn orcid.org/0000-0001-8912-4840, DePaolis, Rory A and Keren-Portnoy, Tamar orcid.org/0000-0002-7258-2404 (2009) A dynamic systems approach to babbling and words. In: Bavin, Edith L., (ed.) The Cambridge Handbook of Child Language. Cambridge Handbooks in Linguisitics . Cambridge University Press , Cambridge , pp. 163-184.

10

A dynamic systems approach to babbling and words

Marilyn M. Vihman
Rory A. DePaolis
Tamar Keren-Portnoy

10.1 Introduction

What is the developmental function of babbling in relation to language, if any? How is it related to the child's first words, and can this relationship shed any light on the highly controversial issue of the origins of grammar in acquisition? Studies of both infant speech perception and early vocal production have produced a wealth of findings over the past thirty-five years, but theoretical progress has been slow, with deductive ideas drawn from linguistic theory often masking the coherent evidence provided by observational and experimental studies.

Dynamic systems theory (Thelen & Smith 1994), with its emphasis on the role of variability in developmental advance, on the independent emergence of related skills as a self-organizing catalyst for behavioural change and on the deep interconnectedness between perception and action and learning, offers a promising perspective on early speech development. While reviewing the empirical findings of studies of production and of links between perception and production this chapter will also consider the relationship of those findings to dynamic systems theory.

10.1.1 The challenge: construction of a first system

A central concern of the study of child language is to account for the developmental source of linguistic knowledge. In one influential approach to this problem innately given Universal Grammar (or UG) is assumed to provide the knowledge of linguistic structure that serves as the starting point for language acquisition, leading to the basic question: *What exactly needs to be learned?* (Peperkamp 2003). This must then be followed by the question of the nature of the triggering process needed to establish the specifics of a given language: *How does the child recognize the critical data that* *will make it possible to set the appropriate parameters, or to rerank constraints in the appropriate way?* (see for example, Fikkert 1994, Lleó & Prinz 1997). For approaches that deny the existence of UG, such as the constructivist approach (see Menn 2006, Tomasello Ch. 5), the questions are the converse: *With what knowledge, if any, does the child begin?*, followed by the complementary question: *How can the child gain knowledge of linguistic structure or system?*

The role of phonology in the development of linguistic knowledge is often given short shrift by researchers interested in word learning (e.g. Bloom 2000, Hollich *et al.* 2000), while production is similarly disregarded by researchers focusing on perceptual advances. Yet before a child can begin to develop linguistic meaning or make referential use of words he or she must be able to represent and access word forms or phrases, which can then come to be associated with recurrent situations, objects or events. Furthermore, it seems shortsighted to assume that perceptual advances alone can suffice to account for language learning. A long tradition of both diary and planned observational studies has found wide individual differences in the rate and pathway of emergence of word production and phonological knowledge across children developing normally, even within the same ambient language group (see Vihman 1996); experimental group studies of word recognition and learning shed little light on this critical aspect of phonological development since it is individuals that learn words, not groups. It is evident that both lexical and phonological learning depend on the development of representations that integrate perception and production; this remains a central issue which has so far attracted insufficient attention.

In this chapter we will adopt the second position identified above, which looks for broad biological foundations to language but posits no specific linguistic knowledge as part of that foundation. Following Braine (1994) we will argue that it is a powerful learning mechanism – coupled with the speech motor system – rather than innate knowledge of linguistic principles that can be identified as the source of the remarkable human capacity for language. Pierrehumbert (2003: 118) proposed that the phonological system is 'initiated bottom-up from surface statistics over the speech stream, but refined using type statistics over the lexicon'. She does not elaborate on the source of the lexical knowledge that supports the second cycle of statistical learning, however. We argue below that the missing link is production experience, which brings the specific adult lexicon to which the child is exposed into focus and into partial or incipient mastery, leading, as Pierrehumbert says, to a new cycle of statistical learning based on types, not tokens. We will seek to show how that learning is first fuelled by the maturational emergence within the first year of vocal production of adult-like syllables. We will demonstrate the role played by babbling practice in supporting attention to and memory for first words, and we will argue that those early words in turn provide a database for distributional learning, the proximal source of emergent phonological systematicity.

10.1.2 Dynamic systems theory (DST) and the origins of grammar

In general, developmental ideas have been scarce in the literature on phonological acquisition, which has tended to draw instead on formal models of adult language and to apply them in a deductive way to child language patterns. Yet when we turn to such a deeply developmental theory as that of Thelen and Smith (1994), we find that their ideas have a remarkable degree of correspondence with the empirical findings which have accumulated over the past thirty-odd years of intensive study of infant speech perception and production, despite the fact that those findings are outside the domain of Thelen and Smith's own research (although Thelen 1991 relates dynamic systems ideas to the development of vocal production).

A key dynamic systems idea is that we must examine *process* in order to understand the origins of structure, which also means accepting *variability* as the very stuff of development. 'In detail … development is messy … What looks like a cohesive, orchestrated process from afar takes on the flavor of a more exploratory, opportunistic, syncretic, and function-driven process in its instantiation' (Thelen & Smith 1994: xvi). In what follows we will first provide a brief account of the process by which babbling is transformed into the first word production.

Nonlinearity is found again and again in empirically grounded accounts of language acquisition as well as in other areas of development. The notion of a predictable succession of categorically distinct 'stages' is generally revealed, on closer analysis, to be a false lead. 'The boundaries of *progressive stages* are … blurred by seeming regressions in performance and losses of previously well-established behaviors' (Thelen & Smith 1994: xvii; our italics). In what follows we will illustrate the nonlinearity of early phonological development, in which the first largely accurate word forms give way to a long period of template-based production, which is less accurate but also more systematic, reflecting the first steps in the construction of a phonological grammar.

According to Thelen and Smith (1994: 247), in a discussion of the emergence of successful reaching for objects in the first year:

> From the messy details of real time – from the variability and context sensitivity of each act – global order can emerge … Knowledge … is not a thing, but a continuous process; not a structure, but an action, embedded in, and derived from, a *history of actions*. (our italics)

In what follows we will attempt to account for the emergence of flexible word-production patterns – different for each child, in accordance with the differences in individual histories of exposure, of 'intake', of early vocal production preferences and of first word use.

10.2 The starting point: biological precursors

Interest in early speech patterns has grown considerably since Jakobson (1941/68) made the claim that babble is wholly unrelated to early word forms, which he took to signal the onset of *linguistic* production. These ideas were shown to be untenable over thirty years ago (Oller *et al.* 1976, Vihman *et al.* 1985); babbling is now generally accepted as providing the raw material for early words. The continuity between babble and first words should not, however, be taken as evidence that the onset of canonical babbling (Oller 1980) is primarily a language-driven activity. There is strong evidence that babble is just one of many rhythmic motor skills that come online in the first year of life, providing the infant with the tools with which to gain knowledge of the world (Iverson *et al.* 2007, Thelen 1981). In Piaget's terms (1952), babble is a kind of 'secondary circular reaction', a perceptuomotor link that helps to lay the foundations for intelligent behaviour.

Campos *et al.* (2000) document the cascading effect of cognitive advances springing from the ability to initiate locomotion. Considered in a social context, the onset of babble can be expected to have a similar cascading effect. Currently there is a growing consensus that babble is best viewed as a multimodal activity, involving both proprioceptive and auditory experience. This provides powerful support for perceptuomotor learning, an excellent illustration of the way that simple linear progression in a basic motor system makes possible the learning of complex cognitive structures (cf., e.g. Rochat 1998, Westermann & Miranda 2004).

The babbling patterns of infants are highly individual and yet subject to very simple biological constraints. The earliest stable supraglottal consonants produced (excluding glides, which are difficult to distinguish from vowels) are stops and nasals (Locke 1983, McCune & Vihman 2001), both of which can be articulated by simple raising and lowering of the jaw. Davis and MacNeilage (1995) have formulated this process in terms of the frame/content theory of early speech organization. In their account, early speech is dominated by successive cycles of mandibular oscillation (the 'frames'), in which the starting tongue position determines both consonant and vowel. Thus, alveolar stops co-occur with front vowels (e.g. [di], velar stops with back vowels (e.g. [ko]), and bilabial stops with central vowels (e.g. [ba]).[1] As babbling becomes more variegated, combining different consonants within a single vocalization, the infant gains control over the 'content' within each syllable, leading to a wider range of consonant/vowel combinations. The co-occurrence of consonants and vowels in early speech has been found to hold in numerous languages (but see Chen & Kent 2005).

[1] For an introduction to phonetics we refer readers to Ladefoged (2006).

The gaining of voluntary motoric control over a specific consonant is the next step toward incorporating these articulatory gestures into early words. McCune and Vihman (2001) tracked these simple early speech patterns – termed vocal motor schemes (VMSs) – in twenty infants. They characterize a VMS as 'a generalized action plan that generates consistent phonetic forms … a formalized pattern of motor activity that does not require heavy cognitive resources to enact' (McCune & Vihman 2001: 152). They operationalized the onset of a VMS as the production of ten or more occurrences of a given consonant in each of three out of four successive 30-minute observational sessions. The VMS thus incorporates an element of both consistency and stability over time. Attainment of a VMS means that the infant is able to consistently access a speech-like motoric pattern with the expenditure of only very limited cognitive resources – freeing those resources to support the novel attentional and memory tasks of associating an arbitrary sound pattern with a meaning.

10.3 The role of babbling: the accuracy of first words, 'preselection' and the 'articulatory filter'

Contrasting their findings with the 'course of phonological development as it has been previously reported' Ferguson and Farwell (1975: 429) noted a number of 'surprising tendencies' in the course of their analysis of the first words of three children acquiring English. The surprises included (a) the relative 'accuracy' of many early child words, with later regression to more primitive forms, (b) the great variability of the early word forms, and finally (c) the 'seeming great selectivity of the child in deciding which words he will try to produce' (Ferguson & Farwell 1975: 429).

The finding of early accuracy has been supported in many subsequent studies (cf. Appendix B in Vihman 1996, which includes the first recorded words of twenty-seven children each acquiring one of seven different languages). To illustrate this, Table 10.1 presents the first four words of a Dutch child, Thomas (based on Elbers & Ton 1985).

Like most early words, the Dutch target words are one or two syllables in length and include mainly early learned consonants (labial and coronal stops, the glide /j/, and /s/, less common in early words but still one of the core consonants in babbling as well as words: See Locke 1983). Somewhat unusually, however, two of the words include two different places of articulation, with a change of both place and manner in *pus*.[2] The child forms are remarkably close to the adult models, if we allow for cluster reduction and a substitution of [x] for /s/ in most forms of /pus(jə)/. Thomas'

[2] Elbers and Ton note that eight of Thomas' first twenty words involved more than one place of articulation; only one violates the sequence front–back seen in *part* and *pus*. This is typical of early melodic patterns: See Jaeger 1997, Vihman and Croft 2007.

Table 10.1. First word forms: relative 'accuracy'

Thomas (Dutch, 15–16 months)		
adult form	*gloss*	*child form*
/auto:/, /o:to:/	'car'	[at], [atə], [aut], [auto:], [o:t], [o:to:]
/hap/, /hapjə/, /hapi/	'a (little) bite'	[ap], [apə], [hap], [hapə], [hab], [habə]
/pa:rt/, /pa:rtjə/	'horse, horsie'	[pa:t], [pa:tə], [ba:t], [ba:tə]
/pus/, /pusjə/	'cat, kitty'	[pusj], [pəx], [bəx], [pux], [bux]

first four words fit the characterization of (more or less) 'accurate'; they are also seemingly 'preselected' for their relatively simple and accessible target forms. Interestingly, Elbers and Ton note that the babbling patterns [at(ə)], [pa:t(ə)] and [bəx], recorded during 'playpen monologues' when the child was alone, 'are already present in babbling *before* their corresponding words are reported to be produced' (1985: 557).

What then is the mechanism underlying the evident 'preselection' of forms to attempt? How can the child know what *not* to attempt? Vihman (1993) proposed that an 'articulatory filter' might be mediating the input, rendering salient those patterns with which the child was already familiar from his or her own babbling production. In this model, the emergence of adult-like syllables, in the middle of the first year, provides the child with a valuable resource (a kind of 'bootstrap', or easily accessible facilitator) for focusing in on selected portions of the fast-moving input speech stream. The tool would be deployed involuntarily: once one or more consonants have been well practised – some weeks or months after canonical babbling begins – the child's attention is likely to be captured by sound patterns that constitute a 'good enough' match to his or her own babbled productions, just as adult attention is sometimes captured by overhearing a highly familiar proper name, for example, embedded in a conversation not consciously attended (Wood & Cowan 1995). By 'good enough' we mean here roughly the same thing as was intended above by 'accurate'. Such an implicit experience of a match of own vocal pattern to input speech would eventually lead to the child's use of such patterns in relevant frequently repeated or routine situations; the consequence would be a small number of known lexical items, the first identifiable words, typically produced only in limited contexts (Vihman & McCune 1994; see Figure 10.1).

A recent experimental study confirmed the existence of something like an 'articulatory filter' by testing the effect of well-practised consonants (VMS) on the child's attention to non-words embedded in short sentences (DePaolis 2006). DePaolis recorded the infants every one or two weeks from 9 to 10 months on and tested them as soon as they had mastered at least one supraglottal consonant to VMS criterion. In order to administer the perception test as soon as the child showed a reliable production

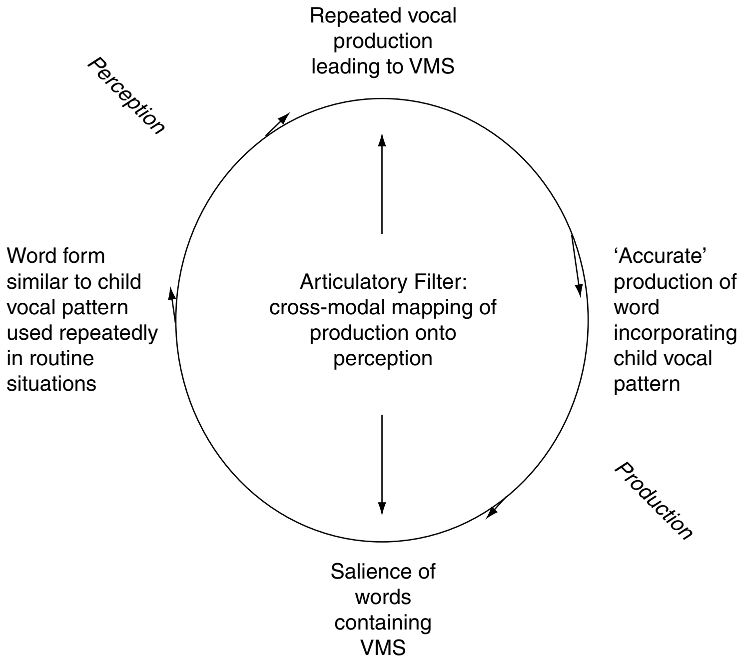

Figure 10.1 The matching of self- and other-produced vocal patterns to own production, supported by a familiar situational and/or verbal context, helps the infant to 'choose' relatively accurate first words.

preference, VMS was defined operationally either as in McCune and Vihman (2001, see section 10.2), or, alternatively, as fifty or more occurrences in the course of one to three sessions.[3] Testing involved presentation of three types of brief contrasting passages of five sentences, each passage consisting of nine uses of non-words featuring (a) the child's VMS (e.g. for /p/b/, *bapeb*), (b) another child's VMS (e.g. for a child producing /t/d/ to less than VMS criterion, *deeted*), or (c) the fricatives /f/v/, which are seldom if ever used to VMS criterion in this period (e.g. *vufev*). The passages consisted of simple sentences with one or two content-word slots filled with the relevant non-word type.

Testing the children within a week of the recording session in which the first VMS was identified proved critical, as the testing revealed a bipolar response to the non-word passages: Of the eighteen children tested, half had only a single VMS; of those nine children, six showed greater attention to the passages featuring their own VMS, while of the nine with multiple VMSs, all but one showed the reverse pattern, greater attention to the 'other-child' VMS passage. Thus, the extent of a child's prior use of a

[3] Voicing differences were disregarded in tallying infant consonant production, both because infants do not control voicing in word production at this age (Macken 1980) and because voicing is difficult to transcribe reliably.

particular consonant had, as predicted, an effect on his or her perceptual attention to that consonant – but the effect shifted from attention to what is familiar to attention to what is novel with the mastery of a second consonant.

Interestingly, production practice has been shown to affect semantic processing as well. In an event-related potential study in which infants heard familiar words that were presented together with (but slightly following) pictures that did or did not match the words, Friedrich (2007) found an 'N400 effect' at 14 months but not at 12 months (see also Friederici Ch. 4).[4] Strikingly, 12 month olds as a group did show an early differential response to the matching vs. the mismatching picture–word pairs (interpreted as a priming effect of the pictures in the case of matching words only), indicating that (most of) the words were recognized when presented in the matching condition. In the mismatch condition conflicting information from picture vs. word was the likely cause of the infants' failure to recognize the words; as a consequence, there was no associated meaning search and no N400 effect. In contrast, a subgroup of 12 month olds with high early word production (five to twenty-nine words) did show the N400 effect, with significantly stronger responses in the children reported to be saying the most words – indicating that these precocious infants were accessing the familiar words and responding with an effort at semantic integration even when the words were out of context in relation to the images they were looking at.

10.4 Word templates: the beginnings of phonological organization

10.4.1 Holistic early word representations: production vs. perception

Early production studies gave rise to the claim that the first phonological representations are whole-word based (Ferguson & Farwell 1975) and 'holistic' or 'schematic' (Waterson 1971). The claim is now controversial, since recent experimental studies, addressing either word recognition or word learning, have seemed to suggest that early (perceptual) representations are, on the contrary, 'finely detailed', giving rise to the 'phonetic specificity' hypothesis (based on eye-tracking: Swingley 2003, Swingley & Aslin 2000, 2002; preferential looking: Bailey & Plunkett 2002; or the 'switch paradigm': Fennel & Werker 2003, Werker *et al.* 2002b). These studies test children's ability to detect differences between novel or familiar words that are minimally distinct phonetically, which involves little or no involvement of prior knowledge, whereas the production studies

[4] In adults, a larger negative deflection (N400) in response to unexpected than expected words in a given context is taken to reflect the effort of semantic integration.

necessarily involve accessing representations in long-term memory, often in the absence of any immediate verbal or situational priming.

The nature of infant 'phonological representation' is as yet poorly understood. Different results are obtained, depending on accentual pattern (English vs. French: Vihman *et al.* 2004) and task demands – specifically, word recognition, word learning and word production. The task differences are important: in the case of word recognition, both the word form and the contextual situation or the image of a referent object may be expected to prime memory for the word and its associations, making the memory load negligible (as in the Swingley and Plunkett studies).

In the case of word learning significant attentional resources must be allocated to the problem of retaining the arbitrary sound–meaning link, as Werker and her colleagues have argued (cf. also Storkel 2001, who made the same point on the basis of a word-learning experiment with 3 year olds). This should make the task of learning new words particularly difficult for children who lack a stock of well-practised production patterns or routines to support memory for the new word form. One indication of this is the finding, reported by Werker *et al.* (2002b), that after habituation training to associate /bɪ/ to one novel object and /dɪ/ to another, the only 14 month olds who responded with surprise to the 'switch trial', in which the new 'word form' is associated with the wrong object, were those with a reported production vocabulary of over twenty-five words (whereas the 17 month olds were 'successful' as a group in showing word learning in this sense). The fact that a larger production vocabulary has been found to be associated with advanced performance as regards both semantic processing of familiar words and novel word learning is a strong indication that production experience supports the accessing and use of familiar word representations (cf. also Mills *et al.* 1997).

The contradiction between the apparently 'detailed' representations suggested by perception experiments and the holistic representations imputed to children on the basis of production studies can be reconciled, then, if we bear in mind that word production requires cognitive resources above and beyond what is required for word recognition or even new word learning – in particular, memory and planning as well as motoric skill. As children begin to make use of larger numbers of word types they must rely on temporarily activated representations for production, often showing regression in accuracy in the word forms they produce. These later representations, although dependent on perceptual experience of a sound pattern, give us good reason to accept Waterson's (1971) judgment that they are holistic 'schemas' or, in our terms, templates, in which the child's previous production practice strongly influences his or her memory for word forms. We will support this contention with examples, below, and will address the question of the source of the holistic representations in our discussion of learning mechanisms.

10.4.2 Whole word phonology: variability

Several arguments for whole word representation as the basis for production are summarized in Vihman and Croft (2007: 689); we review them here, beginning with illustration and discussion of the first, 'variability'. The three remaining arguments – holistic match of child to adult form, similarity among child forms, and response to challenges – will be discussed in the next section.

1. *Variability*: A sound may be produced differently in different early words, and individual words may be more or less variable (Ferguson & Farwell 1975). This suggests that although the child has gained knowledge of particular words ('item learning'), he or she has not yet developed abstract categories of sounds.

Ferguson and Farwell (1975) famously reported twelve widely varying pronunciations of the word *pen* produced in the course of a single session at about 15 months by K, one of the two American children they observed, with alternate production of labial or alveolar, oral or nasal onset, or neither, and with a range of oral or nasal low to mid vowels, as shown in (1):

(1) [mã^{ə} (im.), ʌ̃ (im.), dɛ^{dn}, hɪn, ^{m}bõ, p^{h}ɪn, t^{h}n̩ (x3), bɑ^{h}, d^{h}aʊ$^{\sim n}$, buã][5]

The child K seems to have a holistic auditory image of the word but no clear vocal match for it within her existing repertoire, even with the support of an immediately preceding adult production;[6] the exploratory variation, which seems primarily to target the articulatorily unfamiliar final nasal, clearly reflects the perceptual influence of the final nasal on the word as a whole.

A similar example of a 'hard word', attempted six times by an English child, Jude (also aged 15 months, but already producing twenty-five words in a half-hour session, which corresponds to a cumulative lexicon of over fifty words: Vihman & Miller 1988), is *circle*, variously produced, in full or partial whisper, as:

(2) [tṣɬu, tṣt^{h}ə (x2), t^{h} t^{h}, tɒ̥tɬju̥ (im.), k^{h}tɬʉ̥ (im.)]

Here we see evidence of child attention to the sibilant and its co-occurrence with a stop and a lateral, although the place of the stop appears to be uncertain as does the sequencing of the various segments, again despite the presence of an immediate adult model in two cases. It is evidently not the individual sounds themselves that Jude cannot accurately reproduce,

[5] im. 'imitated'. Note that K had produced no more than eight or nine words in a session spontaneously at this point.

[6] In the full listing of child variants for each word that Ferguson and Farwell included in a later reprint of this paper (1977) we find that K, in the three preceding weekly recording sessions, had produced onset oral and nasal labial stops but only two codas, a weak [^{k}] in [mʌ̃^{k}bu] *monkey* (im.) and [x] in [b$_{w}$ux] *book*. A nasal vowel occurred once, for the first time, in the previous session: [Q̃] *on*, and also in two other words in the current session: [mɑ̃] *me/mine* and [hɪʌ̃], [^{m}kjũ] *thank you*.

since each of them is produced in at least one attempt at the word. Similarly, there is no reason to believe that he cannot perceive the adult segments. Instead, his difficulty appears to derive from the planning and production of the word pattern as a whole, in sequence, with its rapidly changing series of consonantal gestures.

The children's 'underlying representations' cannot easily be inferred from these production efforts. They are better described as dynamic or fleeting than as set or stable (or reliably accessible), with apparent influence on the momentary remembered form of the word not only from the percept of the target word itself but also from coexisting ('whole word') production patterns in the child's repertoire – patterns which must be accessed for vocal expression.

10.4.3 Templates in the word production of three late talkers.

Three further arguments for whole word phonology were cited in Vihman and Croft (2007).

2. *Holistic match of child to adult form*: Comparison of early child words to their adult models on a segment-by-segment basis is often difficult, as Waterson (1971) showed in the case of her son 'P'. Instead, the child appeared to be targeting a 'whole gestalt'.
3. *Similarity among child forms*: The interrelation between the child's own words may be more evident than the relation to the adult models (Macken 1979).
4. *Response to challenges*: The 'gestalts' or 'templates' which are taken to underlie the common patterning of a child's words can be seen as responses to one or more challenges posed by the segmental sequence or structure of the word form as a whole. The primary challenge, in most cases, is the difficulty of producing different consonants, vowels or both within a single syllable of a word (e.g. *pen*) or across syllables (*circle)*.

The relationship of child to adult form and the sources of child difficulty have already been illustrated by the two sets of variable forms presented above for K and Jude, one just beginning to produce words, the other (Jude) having a considerably larger lexicon. Appreciation of the patterning seen in a child's word forms requires that one consider the full set of word forms produced in a given session, however, or over a delimited period of time (e.g. Priestly 1977).

In order to further illustrate these principles and to show their interrelationship we draw here on patterns observed at the 'twenty-five-word point' (25wp: the first half-hour recording session with twenty-five or more words) of each of three British children who were late to begin talking. Similar patterns, templates or 'canonical forms' (Menn 1983) from younger children have been reported in numerous studies, beginning with Waterson (1971) and Menn (1971). For recent crosslinguistic data illustrating template use see Vihman and Kunnari (2006), based on longitudinal observations, and Vihman and Croft (2007), based on diary studies.

Two of the children whose data we present here (Elise and Tony) were identified at 30 months as '(expressive) late talkers' on the basis of having a score within 3 months of chronological age on the Reynell-III Receptive Scale and a score of 6 months or more below chronological age on the Reynell-III Expressive Scale.[7] These children thus differ from the younger children whose data have been presented in illustration of the development of templates in earlier studies by virtue of their larger (age-appropriate) receptive lexicon. It is all the more striking that their limited phonetic resources should result in patterns that resemble those of the younger children. At the same time, their wider ranging lexical targets mean that the 'adaptations' observed are sometimes even more radical than those reported for younger children. The process of induction of templatic patterns that we describe under learning mechanisms, below, can be understood to be the same.

1. Jack (26 months.)[8]

   In this session Jack, who was engaged primarily in 'book reading' with his mother, actually produced fifty-two different word types altogether, excluding word combinations, onomatopoeia and doubtfully identifiable forms. All of the words were produced spontaneously at least once. Two word patterns dominate Jack's production: CVVN, or monosyllables including a diphthong and nasal coda, and CVGlV, or disyllables with a medial glide.

   a. CVVN: Some of these forms are relatively accurate (designated as 'select' in Table 10.2). In each of these 'selected' words the rhyme matches the target, although initial clusters are reduced and the

Table 10.2. Later word forms: the emergence of a CVVN pattern

JACK <CVVN>			
SELECT		ADAPT	
clown	[daʊn]	*boat*	[beɪn]
crane	[heɪ::n]	*ladybird*	[laːbwaʊm]
green	[giːn]	*moon*	[bʊːən]
paint	[beɪn] (x2)	*spoon*	[m̩buːm]
plane	[deɪːin]	*worm*	[beʊm]
train	[dəɪn]		

[7] When first seen, at 25 months, Jack was not yet producing combinations despite having a reported vocabulary of over 100 words on the Oxford CDI (Hamilton *et al*. 2001). At 2;6 he scored within the normal range for both expression and comprehension on the Reynell, however, and so he cannot be considered a 'late talker'.

[8] We discuss the children's word patterns here in order of child age at the 25wp.

Table 10.3. Later word forms: the emergence of a disyllabic CVGlV pattern

Jack <CVGlV>	
Adapt	
banana(s)	[bɛːː | aʊ][9]
bubbles	[bɔːwuːə]
guitar	[giːaː]
Harriett	[heɪjɛː]
pizza	[mbia, biə]
strawberries	[dauːwi]
toast	[dəuːːa]

onset consonant sometimes changes in unexpected or atypical ways (*crane, plane*). In other cases ('adapt') the words show 'adaptation' to the emergent template. For example, two words show consonant harmony (*ladybird, spoon*) and two (*boat, ladybird*) show a change of stop to nasal coda. In two further cases Jack draws out or creates a diphthong: *moon, worm*.

There are three additional CVVC forms with a non-nasal coda. *Plate* [beɪtʰ] seems regular and 'accurate' but does not participate in the pattern; its cooccurrence in the same session with [beɪn] for *boat* shows the unevenness of template use. The remaining two forms have coda [k]: *bike* [maɪʔkʰ] (with its anomalous onset) and *grape(s)* [geɪk], with consonant harmony.

b. CVGlV: In the case of this template there are no 'accurate' or 'selected' productions, although the pattern applies most closely to adult open monosyllables with a long vowel:[10] *bee* [biːa], *no* [nəuːːə], *ski* [ŋiːa], *two* [duːə]. Note that most of these forms also occurred in the same session as monosyllables, CVV_o: *no* [nəuː], *ski* [gi] (x2) and *two* [duː]. The most striking adaptations, however, involve longer words produced with this pattern (Table 10.3). These forms seem to reflect Jack's ease in producing diphthongs, which he can also extend into a second syllable.

2. Elise (33 months.)

Drawing on Elise's 25wp, with 23 imitated and 25 spontaneous words (omitting onomatopoeia), we find a single strong pattern, in which [s] or [ts] are added or substituted for final consonants or clusters.

a. Monosyllables: In the case of monosyllables Elise sometimes seems to be targeting a plural form (*bees, eyes*), but there is reason to doubt that the final *-s* ever has morphological value (cf. *pink, red*:

[9] The vertical line represents a brief pause or break between the two syllables.

[10] All forms are presented here as transcribed; a glide is necessarily present in the disyllabic forms, even where not indicated, as a transition to the final vowel.

Table 10.4. Later word forms: the emergence of a monosyllabic fricative coda pattern

Elise monosyllables <CV(V)s/ts>			
Select		Adapt	
birds	[baːdsː, bɛː[p]s]	*bees*	[weɪːç]
cat(s)	[[t]ɛtsːː]	*bike(s)*	[baɪs]
eyes	[aɪs]	*books* (im.)	[bɪdʔsː]
horse (im.)	[ʰaɪːts]	*cake* (im.)	[kʰiːːʃts]
		cloud (im.)	[waɪːːsːː]
		dog(s)	[dəʔtsː]
		pig	[bɪds] (x4)
		pink (im.)	[bits]
		red (im.)	[weʔðs]
		sheep (im.)	[wɪtsː]
		shoes	[ʒəts]
		socks	[dədsː]
		trees	[wiːːs̪ːː]

im. = imitation

Table 10.5. Later word forms: the emergence of a disyllabic fricative coda pattern

Elise <V_oCVCVs/x/ts>	
Adapt	
ladybird	[əbɛbɛːts]
pirate	[wɛwets]
fairy (im.)	[hɛːwix]
microphone	[həʔdudɛs]
lady	[ɛdiːʃ]
rabbit	[haʔpiːsːː, baʔbiːtsː]
T-shirt	[əʔtɛtʃ]
telescope	[tetətɛːs]

Table 10.4). In addition, Elise produces two monosyllables with coda /n/, *arms* [æːːn] (imitated) and *mouse* [mãn, mãn^{t}, mãn$^{\theta}$] (this may reflect a confusion of *mouse* and *man*, based on a picture involving both a mouse and a pirate).

b. Disyllables. These forms sometimes include the fricative or affricate coda in the first syllable: *cross bones* [dəsbaːn], *icecream* [wisbɹiːː] and even *chicken* [dɪdsən] (with possible metathesis of the sibilant release of the onset affricate), all imitated. More often the coda is in word-final position, for both vowel- and consonant-final word targets (see Table 10.5).

Elise's remaining disyllabic forms with codas have either /m/ (*balloon* [ələuːm] or /t/ (*boat* [bəʔat^h], *pepper pig* [haʔbɛbɪt^h, both imitated). Interestingly, although Elise sometimes inserts a final [s] where none is warranted, she never omits a coda altogether when the target has one.

3. Tony (35 months.)

Tony, the latest of the three children to reach the 25wp (when he produced 33 different words spontaneously), has a dominant word pattern <v$_o$CVv$_o$>, the largest subset of which shows the more specific pattern <v$_o$wVv$_o$>. In both cases Tony tends to add a filler [(h)V] before the word if there is none in the target.

a. Stop or nasal: In the case of words *not* produced with medial [w], labial and velar stops and nasals occur initially or medially (Table 10.6); in the case of two target words with /f/ onset Tony produces anomalous substitutions (*fly, four*) – in both cases using an output pattern that serves elsewhere for a 'selected' word (*bye, go*). There is also one disyllabic target adapted for production with reduplication of the velar-onset first syllable ('*copter* [gÅ/gÅ/] (x2)), which is again similar to a frequent output syllable (cf. *(a) car, all gone* as well as *go*).

b. Medial <w>. This more specific pattern is produced as a match to target ('selected') in five words or phrases, while in ten additional words Tony imposes the pattern, sometimes at the expense of quite radical changes to the target word form (e.g. *carry, soil*: Table 10.7). In addition, two words are adapted to this template but include a (harmonizing) labial coda: *bum* [awʌm], *Tom* [əwɑːːm]. Tony produces codas in only three other words, all monosyllabic targets; all harmonize coda with onset: *beep* [biːp^h], *dig* [hɛgɪg] and *stuck* [gɒk^h, ɒgʊk^h]. It is striking that Tony uses no coronal consonants at all.

Table 10.6. Later word forms: the emergence of a <VCV> pattern

Tony <V$_o$CVV$_o$>			
Select		Adapt	
(a) ball	[ɒ:bɔ:] (x4)	*please*	[heɪ: bi:]
(a) bike	[æʔba:]	*train*	[ɒgeɪ:::]
bye	[baɪ:]	*fly*	[əbaɪ:]
(a) car	[hæga:, a:ga:] (x2)	*four*	[əgɔ::]
all gone	[ɔ:gɒ]		
go	[gəu::]		
(oh) no	[ŋəu:: (x3), ɔ:əŋəʊ]		
more	[mɔ:] (x3)		

Table 10.7. Later word forms: the emergence of a <VwV> pattern

Tony <V_owVV_o>			
Select		Adapt	
all wet	[aː wɛʔ]	*aeroplane*	[aʊwɛ]
away	[aweɪː]	*carry*	[əwiə]
hurray	[həweɪː]	*flowers*	[aːwe]
wee	[wiːː] (x2)	*fly*	[ɒʔwaɪːː]
whoa	[wəuː]	*over*	[əuːwɛ]
		soil	[hawaʊ, əwaʊː]
		that way	[ɒ.weɪː]
		up there	[ʌʔbwɛː, aːbwɛː]
		wheelbarrow	[aʔwɛː, awɛː]
		wire	[əːwaːː, ɛwa]

Alongside his strong labial bias, expressed in his ‘choice’ or discovery of <w> as a template consonant, he also produces many words with [g] and substitutes a velar nasal in the word *no*.

The patterns we see in the words produced by these three late talkers reflect, as do the patterns of younger children, their reliance on a small core consonant inventory, one which primarily consists of stops, nasals and glides. Beyond that, we see in the many ‘adapted’ forms, or forms which fail to match the target (even in cases where the child clearly has the necessary articulatory or phonetic resources to make a more accurate match, e.g. Jack’s *boat, toast*), evidence that the children are inducing generalized patterns from their own output. That is, once the child has learned a certain number of adult-based words, usually at the fairly slow pace characteristic of ‘item learning’, word learning becomes easier (as evidenced by a rapid increase in new word production). This greater facility can be ascribed to the emergence of one or more well-practised ‘motor plans’ or templates that serve to support attention and memory to the form–meaning link. We see this as the beginning of phonological systematicity – in other words, as an emergent phonological grammar, in which the child goes beyond individual word forms to develop patterns representing possible word shapes which are based on the intersect between his or her own output forms and common input patterns.

10.5 Learning mechanisms

Studies of artificial grammar learning in adults (e.g. Reber 1967) already suggested the importance of statistical or ‘distributional’ learning over forty years ago, but it is only in the past decade that experimental findings have made it clear that children, like adults, automatically tally distributional regularities in the environment (Saffran *et al.* 1996a; also see Thiessen Ch. 3). This learning capacity is not restricted to speech (i.e. is not 'domain specific'), however, but has been shown to apply automatically to any regularly recurring sequence in the infants' environment (Kirkham *et al.* 2002). If we relate these findings to the host of experimental studies of prelinguistic responses to speech reported in the 1990s (Jusczyk 1997), we can conclude that over the course of the first year infants gradually gain a sense of input language patterning as regards sequences at any level of linguistic organization – segments, syllables, accentual patterns, words, phrases, clauses. Based on adult studies (e.g. Saffran *et al.* 1997), it is clear that this learning occurs in the absence of any specific intent to learn or even of (conscious or focused) attention to linguistic patterning as such.

However, word *production* requires that the child register arbitrary form–meaning relationships; the word forms repeatedly used in a given situation must persist in the child's memory, together with their context of use (or meaning), in order to lead to recognizable word use. This need not imply conscious attention or a specific intention to learn. Rather, the routine recurrence in a given situation of a sound pattern familiar from the child's own vocal practice can be taken to prime the child to produce that pattern in the often experienced situation (see Fig. 10.1). Each such use – which necessarily involves motoric effort (Elbers & Wijnen 1992) – can be expected to strengthen the memory trace, making future deployment of the same pattern more likely (Edelman 1987) and supporting memory for both form and meaning. Such early word production, supported by the experience of a perceptual match, can be taken to be the source of the relatively 'accurate' first words, as indicated above. This is 'item learning'; each word must be remembered individually as a whole, form and meaning together. It is thus quite different from the rapid, automatic registering of recurrent regularities ('distributional learning').

Current thinking in neuroscience supports the idea of a dual memory system. It is widely accepted that the hippocampus is required to consolidate detailed, multimodal episodic memories, which are the basis of learning from unique experiences, such as the item learning just described (McClelland *et al.* 1995, Squire & Kandel 1999). Furthermore, the registering and recall of arbitrary form–meaning pairs also generally depends on processing in the frontal lobes (known to be involved in the selection of percepts for focused attention). In contrast, the registration of regularities – the essence of distributional learning – occurs even in the face of hippocampal damage, permitting amnesic patients to abstract structure from a set of related items, for example (Knowlton & Squire 1993).

There is thus ample evidence to support a distinction between two types of learning – one probabilistic, statistical, sensitive to distributional properties such as frequency of occurrence and sequential patterning, the other responding to chance conjunctions of unrelated elements (notably, for our purposes, the arbitrary association of form and meaning), essential for the construction of a lexicon. What is most important is the idea that once motor production begins to highlight words in the input, leading to item learning, the 'input' to the child's distributional learning mechanism will necessarily begin to include the child's own word forms. This is a critical change: now the internal structure of the first words – the 'selected' target words, as (1) filtered through the child's primitive speech production mechanism and (2) analysed through distributional learning – will automatically be induced, providing the child with implicit phonological patterns that can be 'projected' onto the input speech stream, 'capturing' possible words to say which will gradually become more ambitious, less close to the vocal patterns actually available to the child. The new words need only share a minimal resemblance to the induced patterns and will be altered in individual ways, resulting in templates such as those described here.

The whole process is data-driven from the bottom up and self-organized through the powerful learning mechanisms highlighted above. Furthermore, at the same time that the infant is producing new word forms that conform to an internally developing templatic system, he or she is also gradually moving closer to the adult system through ongoing implicit comparison of child to adult word forms. As suggested by Pierrehumbert (2003), who supposed that the process happens only much later than the period of the first words, once the child has a much larger lexicon, 'type statistics' can be induced from his or her internal word representations, creating more or less well-defined templates and greatly facilitating and accelerating the process of further lexical learning.

10.6 Conclusion. From babble to words: a developmental account

In order to better understand the processes that might account for the origins of phonological system we have presented some of the evidence to support the essential continuity between babbling and first words. We also claimed that babbling is only one of many manifestations of the child's general motoric development, with its rhythmic base and its cascading socio-cognitive consequences. And we argued that a child's babbling practice provides the essential resources for the identification and shaping of early word forms. We provided experimental evidence to back up the claim that the apparent preselection of adult targets reflects implicit multimodal matching of the child's own vocal production patterns to frequent input speech sequences. In dynamic systems terms, maturational advances in vocal production – primarily the emergence of rhythmic canonical babbling syllables in the middle of the first year – provide fuel for a phase-shift to first word production. But the presence of speech-like syllables in repertoire is not in itself sufficient to catalyze this shift. Instead, the normal environment of a growing child – the presence of talking caretakers, the infant's sense of reward elicited by the production of vocal forms that echo some of that talk, the proprioceptive feedback obtained from the articulation of the syllables which provide that reward – makes available numerous supporting experiences to tune those syllables in the direction of the ambient language and eventually to register, in the child's mind, matching input sequences along with their situational context or meaning (see also McCune 1992).

The route from babbling to words that we described is 'universal' but also highly individual, since the starting points (the particular first syllables or consonants to be mastered) differ as do the pathways followed. We noted that particularly challenging word forms may give rise to an exceptional degree of variability (for evidence of an increase in the variability of a child's word forms in the weeks immediately *preceding* the first manifestation of a stable templatic pattern see Vihman & Velleman 1989, Vihman *et al.* 1994). We also considered both first words (Table 10.1) and later words (three late talkers). In all cases we saw individual phonetic constraints deriving from variable motor skills and practice and we saw that those constraints translated into particular pathways leading to phonological structure. Non-linearity was reflected, if indirectly, in the late-talker word patterns, in which the 'adapted' word forms were sometimes quite remote from their targets yet close to many other forms produced by the child. As outlined by Thelen and Smith, knowledge here again reflects the history of actions of each child, although we did not here trace individual babbling patterns through the accurate first words to the generalized patterns of the later words. We did see that the children construct knowledge each in their own way, based on their own specific perceptuomotor experiences. Finally, we argued that there is no need to posit innate knowledge structures (UG) in order to explain the emergence of language. The learning mechanisms we invoke, unique in humans due to the combinatory power of distributional and item learning, seem to us to be sufficient to account for the formation of a phonological system.

Suggestions for further reading

Ferguson, C. A. & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419–439. Reprinted with Appendix in W. S.-Y. Wang (Ed.). (1977). *The Lexicon in Phonological Change* (pp. 7–68). The Hague: Mouton.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115–154.

Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Vihman, M. M. (1996). *Phonological Development*. Oxford: Basil Blackwell.

Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics*, 45, 683–725.