Chapter 1

Word learning and the origins of phonological systems

Marilyn M. Vihman, University of York

Introduction

Only a few years ago many people presumed that learning language required explicit attention. The phenomenon of 'joint attention' (of caretaker and child to an object of interest) is widely accepted as constituting at least one of the foundations of language development (Tomasello 2003), and evidence that attention in early infancy can predict later levels of intelligence or word learning has also been reported (e.g., Ruddy and Bornstein 1982). However, recent experimental work with adults, infants and children has provided strong evidence of the power and importance of implicit learning, or learning that takes place outside of attentional focus (Gomez and Gerken, 1999).

One place to look for the role of implicit learning is at the earliest stages of word learning, when both speech perception and vocal production are beginning to be shaped by the ambient language, but the arbitrary form-meaning associations that underlie referential word learning have not yet begun to be established. At this stage the infant is beginning to use just a few familiar words or phrases in familiar contexts. We can follow the shaping of infant speech from this early stage, through to the targeting and production of selected adult words, and on to the period of expansion that follows, and consider the role of implicit learning as the child advances into the adult phonological system.

Learning mechanisms

The empiricist vs. nativist debate led Braine (1994) to conclude that accounting for language learning *without* positing specifically linguistic innate knowledge (i.e., without Universal Grammar) would require that one posit instead powerful 'innate learning mechanisms'. One such mechanism is statistical or distributional learning (Saffran, Aslin and Newport 1996).

Distributional or statistical (probabilistic, 'implicit' or 'procedural') learning

The conceptual distinction between explicit and implicit learning is not new (Reber 1967), but only in the past few years have experimental findings made it clear that both children and adults automatically tally distributional regularities to which they are incidentally exposed while attending to a completely different task (Saffran, Newport, Aslin, Tunick and Barrueco 1997). Implicit learning has been shown to occur even in infants when they are exposed to uninterrupted sequences of syllables lacking any natural speech prosody (Saffran et al. 1996). These studies reveal probabilistic (statistical, distributional) rather than categorical learning or 'symbol manipulation'.

If we relate these findings to other experimental studies of prelinguistic responses to speech (reviewed in Jusczyk 1997), we can conclude that infants gradually gain a sense of the patterns in the ambient input

language at the level of segments, syllables, accentual patterns, words, phrases, and clauses, without any intention to learn. In fact, these studies demonstrate a powerful capacity for distributional learning, including a capacity to 'pick up' the linguistic patterns in input speech. Moreover, this learning capacity seems to be a general one, applying to any regularly recurring sequence (aural, visual, tactile, etc.) in the infants' environment (Kirkham, Slemmer and Johnson, 2002). It seems, in other words, to correspond to Braine's idea of a 'powerful learning mechanism'.

Studies of infant responses to speech have revealed emergent sensitivity to what is known as "prosodic coherence" in ever smaller prosodic units over the course of the first year. That is, infants listen longer to sequences that reflect natural as compared to unnaturally interrupted prosodic units: clauses, in infant directed speech, as early as four-and-a-half months, then phrases at nine months and finally words at eleven months (Jusczyk 1997). It is highly likely that these effects are achieved through implicit learning, because although this kind of learning is not strictly speaking 'statistical' or 'distributional', it is not arbitrary, symbolic, or based on attention either. It can safely be termed implicit learning, or learning in the absence of voluntary or focussed attention to the stimuli, intention to learn or conscious awareness of learning.[1]

The effect of implicit perceptual learning can also be seen as infants shift the production of their vowels in babbling towards those in the language(s) to which they are exposed (Boysson-Bardies, Hallé, Sagart and Durand 1989). Similar production effects are reported for prosody

(rising pitch is more common in the babbling of French than in that of American infants in the age range 6-12 months (Whalen, Levitt and Wang 1991)) and for consonants (a larger proportion of labials are found in the vocalizations of 10-month-olds exposed to English and French than in those of infants exposed to Japanese or Swedish (Boysson-Bardies and Vihman 1991)). In each case the differences mirror those in the relevant adult languages.

An instructive test of the idea of implicit (or frequency-based) vs. explicit (or lexically based) learning can be seen in a comparison between infants exposed to Finnish and infants exposed to Welsh. Finnish has a systematic distinction between long (geminate) and short consonants. Welsh does not have two different types of consonants, but rather lengthens consonants between vowels under accent. The result is that children exposed to Welsh tend to hear more long consonants than those exposed to Finnish, since in Welsh they occur regularly in running speech whereas in Finnish they occur in only about a third of the content words mothers use with their children. The median length in Welsh mothers' child-directed disyllables is 118 milliseconds vs. 75milliseconds for Finnish, a highly significant difference. As predicted by implicit learning Welsh infants produce longer consonants, on average, in babble and first words than do Finnish infants (170 milliseconds vs. 116 milliseconds). However, by the time the children reach the end of the single word period the situation has reversed, as Finnish children have learned enough words to have picked up on the occurrence of long consonants. Direct measurement reveals that Finnish medial consonants attain a mean of

223 milliseconds at this point, while Welsh children show little change from the earlier stage (Vihman 2001, Vihman and Kunnari 2006).

Implicit pattern learning has a less direct effect on production, because production is only a secondary reflection of the child's perception of adult speech. To account for an effect on production we must assume that infants are biased to *reproduce* more often the vocalizations they perceive as better matches to what they are *hearing* with greatest frequency in input speech. If this interpretation is correct, the effect should be seen only in patterns that have a solid grounding in the infants' production repertoire, i.e., that are also *produced* with sufficient frequency to be subject to such a 'pruning' effect of the perception/production match. This seems to be true. Phonetic categories such as coda consonants (consonants that end syllables), which are rare in infant babbling, show the impact of distributional learning at a later stage than phonetic categories that come under infant control earlier, such as the long consonants between vowels mentioned above (Vihman and Boysson-Bardies, 1994).

In fact, ease of production and input frequency interact. Children exposed to English and French, on the one hand, and Finnish, on the other, show similar ranges of medial consonant duration at the earliest stages of word production, with some children in each group producing far longer consonants than are typical of adult English or French (Vihman and Velleman, 2000). The Welsh children followed in the study summarized above showed the effect of adult long consonant production already at

this early stage, reflecting *both* the relative ease of production and the relatively high input frequency of this phonetic feature in Welsh. As word learning progresses in each group, infants exposed to English and French restrict the length of their medial consonants in accordance with adult norms while Finnish children increase theirs. The Welsh children show no further lengthening, since in their case lexical learning does not increase the salience of the (purely phonetic) long medial consonants in the input.

In contrast, by the end of the single word period Finnish children are producing even more long consonants, proportionately, than are found in input speech: a mean of 47% of their words are transcribed as including a long consonant, even though only 38% of their words have long consonants as produced by adults. It seems to be that once one or two of the common words with geminates (such as *anna* 'give', *kukka* 'flower', *loppu* 'finished, all done', *pallo* 'ball', *tyttö* 'girl') have been produced, practice leads children to focus on words with long medial consonants in Finnish despite their relatively low overall frequency in the input. This reflects an effect of 'top-down' (cognitive, lexical) processing in contrast to the 'bottom-up' or purely 'signal-based' salience of long medial consonants in a language like Welsh, where they occur more frequently due to their status as phonetic markers of accent. I return to this issue below.

Lexical or symbolic (categorical) learning ('explicit' or 'declarative')

Lexical learning is not the same as implicit learning, however, because word learning must depend at least in part on infant attention. Specifically, the first 'true' referential or symbolic word is demonstrated when a child either comprehends or produces an adult-based word pattern in a new situation (Bates, Benigni, Camaioni and Volterra 1979, Vihman and McCune 1994). Before that can happen, however, the child must have attended to adult use of the word in a situation in which both form and use were clear and salient. Indeed, before word use becomes well established children have been shown to spontaneously deploy such markers of attention as pointing, 'showing', and grunting, all good indicators of intentional communication (Vihman 1996).

By the first half of the second year infants typically demonstrate an interest in language and an intention to learn as well as a capacity for explicit attention to, and memory for, word use. Furthermore, in mothers' speech to their one-year-old infants, words used repeatedly in isolation – that is, words which can most easily capture the child's attention – have been shown to correlate significantly with first word production (Ninio 1992, Brent and Siskind 2001). Thus, despite the well-established abilities of prelinguistic infants to make implicit use of phonetic cues in the input as to how to segment the speech stream (Jusczyk, 1997), words or phrases that are readily available to the child's attention *without* the need for segmentation of the speech stream (because they have been heard in isolation) seem to be the most readily incorporated into the emerging *production* lexicon. It is plausible, then, to make a distinction between learning with and without attention: This would correspond roughly to the

distinction between explicit and implicit learning in adults (although even in adults it is difficult to make a sharp or categorical distinction between the two types of learning or memory (Jacoby 1991)).

Current thinking in neuroscience supports the idea of a dual memory system along implicit vs. explicit lines. It is now widely accepted, based on studies of amnesia as well as on animal studies (e.g., Squire 1992, Baddeley et al. 2001), that the hippocampus is involved in consolidating detailed, multimodal episodic memories, which are the basis of any one-off learning, or learning from unique experiences. Furthermore, the registering and recall of arbitrary form-meaning pairs depends on processing in both the frontal lobes (known to be involved in the selection of percepts for focussed attention) and the hippocampus (McClelland, McNaughton and O'Reilly 1995; for a review, see Ellis 2005). In contrast, the registration of regularities – the essence of distributional learning – occurs even in the face of hippocampal damage (Knowlton and Squire, 1993). Slow skill learning (based on sufficient practice) also occurs without involvement of the hippocampus (e.g., Wilson, Maruff and Lum 2003), as does the gradual learning of repeatedly occuring perceptual consistencies (e.g., Willingham, Nissen and Bullemer 1989). Based on connectionist modelling, McClelland et al. argue that two distinct learning or memory systems evolved for a good reason, namely, to permit rapidly learned novel patterns to be added to an existing system without catastrophic interference (p. 432f.).

Integrating distributional with lexical learning: The spiral model

The evidence thus solidly supports a distinction between two types of learning. But what are the implications for language development? It seems plausible to distinguish three types of learning for first language acquisition. Two of these are the product of 'implicit' processing by the 'non-declarative' or 'procedural' system (Squire and Kandel 1999), and the third requires processing by the declarative system.[2] The three types of learning are the following:

(1) procedural tallying of *regularities perceived in sensory data* of any kind (e.g., of frequencies of occurrence, sequencing, or rhythmic patterning),

(2) declarative (categorical, symbolic) registering of arbitrary form-meaning co-occurrences or associations, leading to a *mental lexicon of linguistic items*, and

(3) 'secondary' procedural induction (a kind of 'tallying') of the *regularities inherent in the linguistic items* registered in the mental lexicon, ultimately resulting in *abstract knowledge of the linguistic system* – the basis for the purely language-internal components of phonology and morpho-syntax.

Type 1 learning is sufficient to account for the advances in language development that have been reported for the first year of life. This kind of learning requires no pre-established 'knowledge base' (Murphy, McKone and Slee 2003) and can thus begin to 'inform the child about the world' just as soon as the infant's sensory organs are complete. As the child gains increasing knowledge, first of prosody (especially rhythmic patterning: Nazzi, Jusczyk and Johnson, 2000), then of segmental

sequences, this type of learning will yield a more detailed analysis of the ambient language.

The second memory system and thus the second type of processing and learning typically come 'on-line' only during the first half of the second year. It makes possible the declarative learning of linguistic units, or referential (symbolic, generalized) word learning. Once children have achieved the 'nominal insight' (i.e., that individual word tokens or exemplars refer to word categories or 'types'), each new encounter with a given word form in an identifiably related situation is taken to belong to the same word type or 'lexical category' as on the previous occasion. For example, the live dog barking next door can be referred to using the same word form as the stylized doggy found in the picture book or on the side of a cup.

This understanding, and thus this kind of access to the mental representation of experiences, cannot be expected to emerge until a stable base of frequently heard words or phrases has developed, along with a capacity for rapidly retaining both phonetic and semantic representations. This is necessary to free up the attentional resources required for declarative ('explicit' or conjunctive) learning: Attention is needed to enable the child to relate new forms to new referents (Werker, Fennell, Corcoran and Stager 2002). Once such voluntary access to lexical representations becomes possible, a lexical knowledge base will begin to be established, somewhat different for each child at first (because

experiences differ), but converging over a period of years on a lexicon very similar to that of other members of the same speech community.

Finally, once a child has begun to establish a lexicon of words or phrases with both phonological form and semantic content, 'secondary' procedural or distributional learning will automatically occur, as the neocortex again goes to work on recurrent regularities (type 3 learning). The input to that implicit or procedural processing is now no longer at the level of direct perceptual input but is rather a representational derivative of the cognitive processing that created the lexical entries – hence the term 'secondary'. As Karmiloff-Smith (1992) has put it,

> 'a specifically human way to gain knowledge is for the mind to exploit internally the information that it has already stored…, by re-describing its representations or, more precisely, by iteratively re-representing in different representational formats what its internal representations represent' (p. 15).

Beyond that difference in 'raw material' or input to the processor, however, the learning process itself must be the same as in type 1. This secondary procedural learning can now be understood as gradually building up the abstract knowledge of system or structure to which we generally apply the term 'grammar' (Pierrehumbert 2003). Importantly, this kind of pattern induction is a good candidate for a pattern recognition system able to account for language learning without the need for innate 'foreknowledge' of linguistic structure in the form of Universal Grammar.

In the discussion so far we have elaborated on and illustrated a conception of phonological development that assumes that 'primary' (1, above) and 'secondary' (3, above) procedural learning are separate contributions, with declarative learning (2) serving as the mediator between the two. Figure 1 – adapted from Vihman and Kunnari, 2006 – illustrates this conceptualization in the form of a 'spiral model', by which procedural or implicit learning 'sets the stage' while declarative or explicit learning adds concrete lexical items to the mix. Once the process has functioned repeatedly to this point, yielding a small lexicon, procedural or implicit memory is triggered again, resulting in new levels of phonological knowledge. The process may be supposed to function continuously over the life-span, although new lexical learning becomes less frequent in the native language once an adult-like level has been achieved, typically late in the teenage years.

INSERT FIGURE 1 ABOUT HERE.

Early words and the link between perception and production

It is likely that first word use, which is typically limited to routine contexts, reflects the implicit matching of vocal production to perceived word forms. I first elaborated this hypothesis in Vihman (1993), although the basic idea had been hinted at in earlier work by Locke (1986) and Kuhl and Meltzoff (1988). I used the expression 'articulatory filter' to convey the idea that the child's familiarity with his or her own vocal production patterns makes sequences in input speech that are like those patterns particularly salient. This idea is in line with a range of other evidence that there is continuity between babbling and speech, both in the general patterning of babble in relation to early word production

(Oller, Wieman, Doyle and Ross, 1976) and in the particular babbling of individual children in relation to their own first word forms (Vihman, Macken, Miller, Simmons and Miller 1985). Studies have found that even the first signed words of children learning a signed language are rooted in prelinguistic gesture (Cheek, Cormier, Repp and Meier 2001).

At this developmental point the range of different word forms in terms of length in syllables, syllable shapes and consonant and vowel types is limited and is broadly the same regardless of ambient language. However, within those limits, individual children learning the same language differ in vocal patterns and draw their particular word forms from their own personal repertoire. That is, each child's early word patterns can be traced to that particular child's vocal practice or babbling.

The apparently nonlinear advance in the accuracy of first words has been a topic of interest for several decades. Over thirty years ago Ferguson and Farwell (1975) suggested, on the basis of three one-year-olds learning English, that the very first identifiable words children produce are relatively accurate. After the rather slow initial build-up of new words, however, there seems to be an overall reduction in accuracy accompanied by an increase in systematicity or inner coherence among the child's own forms (the so-called U-shaped developmental curve) – as well as a more or less abrupt shift to more rapic lexical learning.

Table 1 provides examples of early words from four children each acquiring one of three languages: British English, Estonian and Italian.

The English and Italian children have been selected from a larger group of longitudinally recorded children (12 and 11, respectively), to illustrate relatively rapid and relatively slow first word production.[3] We will return to an analysis of the 'accuracy' of these first words below.

INSERT TABLE 1 ABOUT HERE.

As well as being (relatively) accurate first words have been found to be based on a highly restricted set of adult forms, so that the limitations of the child's early phonology is already in evidence from the words targetted. This finding aroused surprise and scepticism when first reported by Ferguson, Peizer and Weeks (1973), but it was later tested and found to be supported by experimental work (e.g., Schwartz and Leonard 1982) and the finding is no longer in dispute.

What are the implications of the evident continuity between babbling and early word forms? What is the mechanism behind 'accurate' pattern production and word selection based on sound pattern? Once children are able to combine rhythmic jaw movement with voice to produce 'canonical' babbling, or CV-CV vocalizations with adult-like timing (Oller 2000, Davis and MacNeilage 2000), they have access to the first truly adult-like production patterns. From this point on they can begin to gain 'inside knowledge' of what it feels like to produce particular sound patterns in familiar situations. That is, the child will implicitly 'learn' both the 'feel' and the 'sound' of the vocal patterns that he or she produces the most consistently. This idea, that the child may be 'experiencing the flow of adult speech through an "articulatory filter" which selectively enhances

motoric recall of phonetically accessible words' (Vihman, 1996, p. 142), is based on the following logic:

1. An (implicit) *match* of child vocal pattern to adult speech can provide the infant with the 'inside knowledge' referred to above through distributional learning of his or her own repeated articulatory production.
2. The 'selection' evidenced in accurate first word forms must result from *including known patterns in the selection*, not from excluding what is not known. That is, what is familiar becomes salient in the input and is registered or remembered more robustly than what is unfamiliar (Fennell and Werker, 2003).

We turn now to experimental evidence in support of the articulatory filter.

Experimental evidence

My colleagues and I recently undertook a series of studies designed to test the articulatory filter hypothesis. To do this we used the notion of 'vocal motor scheme' (VMS), based on earlier work by McCune and Vihman (2001). A VMS is '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 and Vihman 2001:152). In practice, a VMS was a consonant used 10 or more times in at least three out of four longitudinal sessions (based on monthly recording and transcription). Vihman and Nakai (2003) recorded infants bimonthly, from ten-and-a-half to twelve months of age, and tested them for a perception effect two weeks later, at twelve-and-a-half months, using the Head Turn Preference Procedure (Kemler-Nelson et al., 1995).

Based on data from 27 children acquiring English and 26 children acquiring Welsh in North Wales a small overall effect of production on perception was found. Specifically, nonwords, including a glottal or glide (taken to be precanonical and thus neutral in relation to the issue of production practice) and either /t/ or /s/ for English and either /b/ or /g/ for Welsh (the two members of each pair being roughly equal in input frequency) elicited a novelty response, such that infants who were producing /t/d/ tended to look longer in response to /s/ while infants who were producing either /p/b/ or /k/g/ tended to look longer in response to the stop that they were not yet producing.[4] A subsequent reanalysis showed that the effect, at least in English, was driven largely by the response of children who between them produced over 200 tokens of the VMS consonant across the four recording sessions (/t/d/ in all cases for English; /s/ being relatively rarely produced at this age). Thus, we found that the extent of a child's use of a particular consonant in production did, as hypothesized, affect the level of the child's perceptual attention to that consonant: A consistently produced consonant elicited less attention than a rarely produced consonant.

DePaolis (2006) undertook a finer grained longitudinal study, recording the infants every one or two weeks from the age of nine or ten months and testing them as soon as they appeared to master at least one consonant to VMS level. In order to be able to administer the perception test as soon as the child showed a reliable production preference, VMS was defined operationally either in the same way as in Vihman and Nakai (2003) or, alternatively, as 50 or more occurrences in the course of one to

three sessions. Testing involved randomly ordered presentations of each of three brief contrasting passages of five sentences each, with nine uses of nonwords 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 nonword type (e.g., for /p/b/: *Wow, my pabep is a buppeb one. Did the bapeb go pubbep below? We pubbep call buppeb a lot. Are your bapeb too pabep over there? I see the bapeb here!*). Differing passages were used for each of the stop consonants. The fricative passage, used as a control for all of the infants, was recorded in three forms to ensure that each child heard three distinct passages, one each for the VMS- and non-VMS stop nonwords and a third for the fricative.

Testing the children within a week of the recording session in which they first produced one or more consonants to VMS level proved critical, as it revealed a bipolar response to the nonword passages: Of the 18 children tested successfully, half had only a single VMS; of those nine children, six showed greater attention to the passages featuring their own VMS stop, while of the nine with multiple VMS, all but one showed the reverse pattern, greater attention to the non-VMS stop passage – reproducing the novelty effect found in Vihman and Nakai (2003).

The differences between the two sets of results appears to be a consequence of the differences in design of the two studies. Firstly, the

perception test in DePaolis (2006) was more challenging that that used in Vihman and Nakai (2003). In the DePaolis study, the child could only identify the presence of a particular consonant after some degree of segmentation of the passage into words, whereas the earlier study provided the easier and evidently less appealing task of attending to words listed in isolation (looking times were shorter in that study). Secondly and most importantly, DePaolis found that infants moved rapidly from practice with a first consonant to practice with one or more others. His 'multiple VMS' sample showed a novelty effect, as had the earlier study, which tested somewhat older children. The novelty effect in the earlier study was also weaker, presumably because it combined the responses of both infants with multiple VMS (who could be expected, in hindsight, to have been more attentive to the novel consonant) and infants with only a single VMS (who would likely have attended longer to the familiar consonant). Given the relatively long delay between the final production session and the perception test in the earlier study it was not possible to determine the exact production status of these infants, but by twelve-and-a-half months we can assume that the majority were producing more than one consonant to VMS criterion.

The implication of both these studies is that production practice affects the way children listen to speech. In fact, the De Paolis study suggests that the shift from attention to the child's *own* VMS at the stage of single-VMS production to attention to *other* VMS at the subsequent stage implies 'a powerful mechanism for segmentation…An infant should be predisposed to segment words that contain sounds that they are producing' (De Paolis

2006:152), an effect that is consistent with the relative phonological accuracy of the first words.

Item learning and 'selected' word forms

Let us consider again the word forms presented in Table 1. Some are, as expected, quite close to the target word forms (those in bold face), while others are less accurate. All of the English children produce over half of their early words relatively accurately; Helena, who makes the latest start, produces all of her first words as good matches. In the case of the other British children the less accurate forms generally involve final consonant omission (as in Jude's *cat,* Ian's *bang, catch,* and Ali's *gone*). Jude unusually targets two longer words, omitting the medial syllable in *Barnaby* and the initial trochee in *caterpillar*. In Estonian the two girls (Virve and Madli) similarly produce more than half of their early words in a way that quite closely matches the target. The two boys make more changes, although the changes largely involve omissions – of whole syllable (Eriku *väike, ammuu*; Raivo *viska, aitäh, banaan*), vowel (e.g., Eriku *tiss, tita, onu*; Raivo *shoe, põmm, vesi*), or consonant (including onset consonant, which is rarely omitted in English: Eriku *suur, väike, vana(ema),* Raivo *viska, hiya*). Finally, the first words of the Italian children are also largely either excellent matches to the target (as in Anna's case) or involve unsystematic consonant or vowel changes (Luca, Nicola) or omissions (Nicola, Nina). In just one case we see consonant harmony, suggesting that Anna is making a systematic change to align the word *caffè* with her other words, which have consonant harmony in the target forms as well as in her production.

In general, the early word sets (the first several recorded words of each child) do not appear to reflect a preexisting template or production pattern (with the possible exception of Anna's). Most of the early word forms of each of the children are fairly distinct from one another, although in Estonian, for example, three out of Virve's first nine words have the diphthong [ai] and both Raivo and Eriku produce disyllables only in the case of target words with an onset labial stop (*pall, banaan*; *päkapikk, paber*). In Italian almost all of the early word forms are disyllabic and both Anna and Nicola favor disyllables with harmonizing adult initial and medial consonants. This appears to be a characteristic of early words in a language with a great many simple reduplicated words in infant-directed speech. In contrast, no one pattern dominates the output for any of the British children. Note that in a comparison of first word forms in English, French and Welsh in relation to the early acquisition of rhythm, Vihman, Nakai and DePaolis (2006) found that the disyllabic words most often used by five children acquiring French were all characterised, like the Italian data here, by simple CVCV structures (with harmonizing consonants in *maman* 'mama'*, papa, poupée* 'doll', but not *chapeau* 'hat') while each of the (American) English children's often used disyllabic words included a diphthong, coda, syllabic consonant, C_1VC_2V sequence, or more than one of these sources of difficulty (*apple, baby, Big Bird, button*).

Returning to Table 1 once again, we can also see that the range of phonetic patterns produced is largely limited in all three languages to one- or two-syllable word forms with stops, nasals, glides and glottals - but

with language-specific differences in the patterns. For example, only in Estonian does [s] occur in all four samples. By contrast, word-onset consonants occur in half or more of the early word forms for most of the children in all three languages, but vowel-onsets or syllabic consonants are prominent for two of the Estonian children only. Use of two different 'true' consonants in a word form is rare overall (i.e., excluding glottals and glides, which are already present in pre-canonical vocalizations). Such a limited range of consonant types and phonotactic structures, cross-linguistically, supports the idea that children are depending on their babbling practice, which is highly similar across differing ambient languages, to lead them into word production. They are also consistent with the hypothesis that 'selection' reflects the child's experience of a match between his or her babbling patterns and often heard, situationally salient input words or short phrases (*who's that, see ya, good girl, oh dear, I see*).

Lexical advance and the emergence of phonological systematicity

What we have observed so far is the product of the implicit matching process, with the articulatory filter resulting in the first (context-limited) word production, due to the priming of a well-practiced child vocal form by the repeated occurrence of everyday situations that bring a semantically fitting and phonologically similar adult word form to mind. This is the period of 'item learning' which we see reflected in the early word forms listed in Table 1. These first words provide, by hypothesis, the data base on which the implicit distributional learning mechanism can

again begin to operate, paving the way to a new level of word learning and an increase, often rapid, in lexical growth. Once a few words have begun to be produced, the child's repertoire and his or her attention to the arbitrary relationship between form and meaning in adult lexical use can both be expected to expand, with wide individual differences in the speed of this advance and in the extent to which it is driven by vocal skill and learning or by independent cognitive or semantic development.

Bates et al. (1979) and Vihman and McCune (1994) have emphasized the shift in infant word learning as the early context-bound words, embedded in daily routines or verbal games, begin to be supplemented by more 'context-flexible' word use, with generalized meanings (although these do not always correspond very precisely to adult usage; see the examples provided by Barrett, 1995). The proposed sequence – first context-limited, then flexible word acquisition – is not uncontroversial (Barrett, 1995; Harris, Yeeles, Chasin and Oakley, 1995; see also the critical analysis of the nature of context-bound early words in Ninio, 1993). And while the disparity in results from different studies may, in part, reflect differences in the definition of the two word types, it is probably sensible to view the distinction itself as gradient rather than categorical. In particular, first flexible or 'referential' words are more likely to be among children's first words if they are children whose comprehension is well in advance of their production. That is, children who begin to talk early typically produce some context-bound words in the earliest period, while children who begin somewhat later may produce words of both kinds from the start (McCune and Vihman 2001).

The first flexible or 'referential' words express children's emerging understanding of the relationship between lexical categories (types) and their uses on particular occasions (tokens) (Vihman, 1996), and typically fall into two categories: nominals, which refer to familiar objects or animals in the child's world (not typically to persons, who tend to be referred to by 'context-bound' proper-name terms unique to specific individuals: *mama, papa, baby*) and 'relational words' (McCune-Nicolich, 1981) or predicates (Deuchar and Vihman, 2005). The latter may take the form (in English) of verb particles (*up/down, in/out, on/off*), expressing path or location, or of forms belonging to a variety of word classes which are used to comment on one or another of the dynamic aspects of events (*bye-bye, all-gone, thank-you, mine, more, no, uh-oh, back),* variously expressing ideas such as occlusion, deictic path, iteration, negation or reversal with the same referential generality that characterizes the symbolic use of nominals. These single-word predicates are harbingers of syntactic advance because they signal readiness to form the longer, more complex structures of the first word combinations (Vihman, 1999; Ninio, 2001; McCune, 2006).

The emergence of phonological systematicity

How does the move to more flexible word use, with referential meanings, relate to emergent phonological knowledge? Logically, there need be no direct relationship between these parallel cognitive and phonological advances, nor is there evidence to suggest a causal relationship (Vihman 1976). Instead, the child's increasing awareness that 'things have names'

leads to more proactive word learning – the 'explicit' learning referred to earlier, with focussed attention and a greater degree of both intentionality and effort. This in turn generally leads to more rapid lexical advance (although this need not take the form of a 'vocabulary spurt' (Ganger and Brent 2004)).

According to one current account, the growing density of lexical neighborhoods (generally interpreted to concern mainly or exclusively receptive vocabulary) provides the pressure that drives holistic word learning to yield to segmental analysis and storage (Walley, 1993; Metsala, 1999). However, doubt has been cast recently on both the extent to which 'holistic word learning' can be supported (based on perception experiments involving mispronunciation of familiar words (e.g., Swingley, 2003), and on the validity of the connection between vocabulary growth and segmental knowledge (e.g., Ballem and Plunkett, 2005).

The idea that whole word representation precedes more detailed, segmental representation was originally proposed by Waterson (1971), drawing on *production* data. It was then taken up and supported by other child phonologists (e.g., Ferguson and Farwell, 1975; Macken, 1979; Menn, 1983); for a recent review of the arguments for this position, see Vihman and Croft (2007). The hypothesis was first tested in relation to word *recognition* with 11-month-old French infants (Hallé and Boysson-Bardies, 1996). In a previous study Hallé and Boysson-Bardies (1994) had shown that by eleven months French infants can recognize words used frequently in the home. The later study tested 11-month-old French

children on their ability to recognize untrained familiar disyllabic words even when the onset consonant of the first or the second syllable had been omitted or changed. Hallé and Boysson-Bardies found that although *omitting* the word-initial consonant effectively masked the familiar (iambic) words (children no longer showed greater attention to familiar than to rare words under that condition), word-initial consonant *change* did not. Change to the medial consonant produced mixed results. In a replication of this experiment with eleven-month-old English children, Vihman, Nakai, DePaolis and Hallé (2004) were able to show that the accentual pattern of the language of exposure was critical: In English, in which the words that children hear the most consistently bear initial syllable stress, change to the onset consonant masked the familiar words whereas change to the medial syllable did not.

There are two important lessons to be learned from these experiments. First, the nature of the language involved is crucial. The onset consonant is particularly critical to word form recognition in languages like English, with predominantly trochaic patterning (Jusczyk, Cutler and Redanz 1993). The impact of the language is not always recognized in the literature, where infants' failure to recognize English words with changed first consonants is often interpreted to mean that they have 'detailed phonological representations' even at seven-and-a-half months (e.g., Jusczyk and Aslin 1995). Secondly, experiments such as those reported in Swingley (2003) involved much older children (19-month-olds – at the opposite end of the steeply rising word comprehension curve documented by Oviatt, 1980). In drawing conclusions regarding developmental

processes considerable care must be taken to match either age with age or milestone with milestone across studies focussing on the same issue.

The analysis presented above suggests that once the child's productive lexicon begins to increase, giving the child repeated opportunities to hear and practice a range of phonetic patterns, distributional learning results in the growth of production routines, to be explored further below. Further experience with lexical units will lead to the reanalysis of the wholes into often-used component parts, yielding familiar phonotactic sequences which themselves support new representational learning of unfamiliar word forms (Storkel, 2001). Production practice – first with babbling, then with word forms – is one source of advances in phonological memory (Fennell and Werker 2003). The product of more efficient word learning is more word production, more production practice, and consequently more implicit pattern learning**.**

Later words: Both 'selected' and 'adapted'

Table 2 provides examples of the phonological pattern-learning exhibited in later word forms, based on data from two of the four children learning each of the languages represented in Table 1. For the children followed longitudinally (i.e., the English and Italian children) the data are taken from the first session in which the child produced at least 25 different word types spontaneously; this typically corresponds to a cumulative lexicon, based on a diary record, of 50-75 words. To match this, the Estonian children's words are taken from the time when they had approximately 50 recorded words.[5]

For each child Table 2 shows a single pattern chosen to illustrate the processes of 'selection' and 'adaptation'. In the case of 'selection' the adult model is reproduced relatively accurately (allowing for some typical immaturity in production – consonant cluster reduction, substitutions of stop for fricative, etc.), while in the case of 'adaptation' the adult target is more radically changed in the set of child forms and the child forms are more similar to one another than to the target. Note that the standard for 'accuracy' has now changed, as the children's phonetic skills have advanced. Coda omission, for example, is not typical of all children and tends to be sufficiently systematic at this point to be considered an 'adaptation' where it occurs. In all cases the child's pattern or 'template' is informally expressed (in angle brackets) to account for the two sets of forms, where the source of the child's 'adapted' forms is at least partially to be found in the 'selected' models which serve as representatives of the adult language to which the child is exposed.

INSERT TABLE 2 ABOUT HERE.

This next step in phonological learning, which we take to reflect the beginnings of phonological organization or systematicity, is illustrated by both the commonalities and the differences between the different children's patterns, both within and across language groups. For English, one of the children with rapid lexical learning, Jude, is contrasted with one of the children who made slower progress, Ali. In Jude's case we see a simple template, the open <CV> syllable, expressed in eight 'selected' forms and seven 'adapted' forms. The selected forms constitute good matches (are 'accurate') except for cluster reduction and, in one case,

vowel change (*blue*). The adapted forms reflect the processes of coda omission or, in one case (*flower*), truncation of the disyllable to a monosyllable. In addition, we see monophthongization (in *mouth)* and, in the case of *cake* and *cheese*, an unusual substitution of [i] for [e] and of [k] for [tʃ], possibly reflecting a single output solution for two problematic forms.

In Ali's case we find more 'selected' than 'adapted' forms for her primary pattern, expressed as the template <(CV)CVC>. The 'selected' words are all closed monosyllables (CVC), although a 'filler syllable' appears to be prefixed in some cases (cf. *bed, crash, feet*). The adapted forms include longer target words, however, and reflect a range of different phonological processes, all of which produce a one- or two-syllable word ending in a closed syllable (labial harmony in *(another) one*, perhaps inspired by the /w/ of *one* /wʌn/, truncation in the case of *biscuit,* metathesis in the case of *Tamar* [tama:r], the name of one of the two observers who recorded the child over a period of several months).

For Estonian, we see phonological patterning of quite different kinds in the case of the two siblings, Virve and Raivo. For Virve we see one relatively 'accurate' or selected form, *banaani,* albeit with a shift in the middle syllable from the repeated low vowel of the target to [i], anticipating the final vowel instead.[6] The adapted forms are all trisyllabic and reflect a bias toward the vowel melody <a…i> (as in *banaani*; remarkably, this bias was apparent already in Virve's first words: See Table 1) or <a…u>. The choice of high vowel for the unstressed syllable appears to follow the

implicit rule, 'use [u] if the target includes [u]; otherwise use [i]'. The last two (unstressed) syllables also show full or partial consonant harmony, with a nasal or /s/ as target consonants (again, as in *banaani*), or the melody <s…n>, specifically in the two cases – *maasikas* and *rosinad* – in which the target word form includes both a nasal and /s/ (regardless of their sequencing in the target word). The word-initial /m/ in *maasikas* is retained and the medial /m/ of *raamatut* moved to word-initial position; again as in *banaani,* only the two unstressed syllables are constrained by the place harmony of the template, which is also observed in the child's more numeous disyllabic word forms (through 22 months: Vihman, 1996, p. 224.)[7]

As in the English cases just described, no one process relating adult to child forms can be identified here. Rather, the overall templatic patterning applies to all of the forms. The extent of 'selection' of potential targets even in the case of the adapted forms is striking: Only words with a high vowel in one of the last two syllables are targetted for trisyllabic production, and only words with either /s/ or a nasal in one of the three syllables. In *lennukit* we also see the imposition of [a] for adult [e] in the first syllable; all of the child's forms have a low or lower-mid back vowel in the stressed syllable. (For a template of similar complexity applied to English, see Priestly (1977).)

Where Virve's template applies only to long words, Raivo (like Jude) shows a template that constrains production to monosyllables, with the further requirement that the monosyllable take the shape CVC with

harmonizing onset and coda consonants. The one selected form (an imitation) is accurate except for the displacement of vowel length to the coda consonant. The seven adapted forms actually involve only three fully distinct output forms: <nVn>, <pap> and <tit> (where the angle brackets indicate abstraction away from the variant phonetic detail of the child's forms). This tendency to settle on a small number of homophonous forms to express a large number of target words was discussed and exemplified in Vihman (1981; cf. also Waterson, 1971). It was a successful strategy in that it led to rapid increase in vocabulary; it also demonstrates the same distributional learning or systematizing that we see in the other children at this stage.

Finally, for Italian, we again have one relatively early word learner and one whose progress is somewhat slower. Anna's selected forms reflect her ability not only to change consonants within a disyllabic form but also to produce medial nasal-stop clusters (although not the onset cluster in *dritte*). The adapted forms all involve truncation, reducing three- and four-syllable words (which are common in Italian basic vocabulary) to disyllabic forms. As in Virve's case, the particular form that the disyllables take depends on the target – especially the rhythmic pattern, with penultimate accent in all of the words targetted – but also on Anna's practiced consonant repertoire. Anna produces [mone] for *maialone,* for example, profiting from her production experience with [m] at word onset but otherwise omitting all but the last two syllables (stressed and post-stressed) to match her disyllabic template.

Nicola's template is one also found in languages like Finnish or Hindi which, like Italian, have medial geminates (Vihman and Croft, 2007): VCV, with omission even of (early learned) onset stops (*cade, pronto*) in the adapted forms. Here again we see the targetting of long words, which are difficult to avoid in Italian; and here again the truncated forms reflect the accentual pattern but also show child sensitivity to the rest of the word, with [are] for *trattore,* for example (see Wijnen, Krikhaar and Den Os, 1994). Note, too, that Nicola preserves clusters, whether heterogeneous or geminates, when they occur within the final (accented) disyllable of the model.

What can we say, then, about the beginnings of phonological system, based on the data that we have examined here? The children's patterns and preferences differ considerably, as expected. For example, some of the children avoid fricatives (e.g., Jude), while others welcome them (e.g., Virve, Ali). Similarly, some prefer to produce monosyllables (Jude, Raivo), while another, Ali, extends some target monosyllables with fillers, and yet another, Virve, is just beginning to produce trisyllabic word forms. Phonologically, there are differences in the complexity of the constraints exhibited in this small sample (in which only one template has been chosen to represent the output of each child), from Jude's minimal CV pattern to Virve's interlacing vowel and consonant harmonies and melodies. There are language-specific characteristics – CVC only for languages in which this pattern is common in the input, for example, as in English and Estonian; VCV for languages (only Italian in the present sample) with geminates or iambic accent. But what the data sets have in

common is that the output patterns do not obviously derive directly from the adult languages. They do not seem to be either directly input-frequency based or closely modelled on the target word forms in all cases. Nor are they the product of a single dominant phonological process or predictable in terms of universal markedness conditions or stages of development (Vihman, in press). Instead, the patterns reflect the individual child's perceptual experience of the input *in relation to* his or her motoric experience or practice – resulting in differences that are as great within as across language groups.

Conclusion

This chapter has focussed on the early stages of word production, first on the transition from babbling to speech and then on the first signs of phonological organization, which can be seen toward the end of the single-word period. It has brought experimental evidence of the effects of production practice on perception to bear on the shaping of first word forms and it has attempted to illustrate 'secondary procedural induction' through the regression in accuracy reflected in the later 'adapted' word forms. Above all, it has sought to indicate that both implicit or procedural and explicit or declarative (voluntary, intentional) learning play a role in phonological and lexical development and it has provided a model in outline of how these two kinds of learning become integrated over the period of single-word production.

In terms of the model presented in Figure 1, this chapter has reviewed some of the evidence for implicit perceptual learning of speech patterns in the first year (Step 1) and how experience of the ambient language also has a direct effect on vocal production in the prelinguistic period (Step 2). The 'articulatory filter' was introduced, which acts to heighten the salience of adult words that constitute at least rough matches to the infant's own vocal patterns (Step 3). First word data (Table 1) demonstrate that this in turn results in (typically context-bound) word production (Step 4) that is relatively 'accurate', different from one child to the next within language groups but broadly the same across children learning different languages. The early unsystematic but relatively accurate word forms contrast with the later words of the same children. Those later words show evidence of regression, with some of them deviating in far more radical ways from the adult targets than did the first words each child produced. In the several-month period that generally separates these later words from the child's first attempts at word production the final steps of the model (Steps 5 and 6) can be expected to have applied repeatedly. These last two steps are placed in parallel as they are not thought to be causally related: Words have now begun to be learned with *attention* and *intention* while the child's productive vocabulary is being implicitly analysed and reanalysed for recurrent patterns, yielding the child-particular word templates illustrated in Table 2.

Overall we see in the first words the value the child gains from being able to produce something that resembles at least the simplest of adult word forms (not only *mama* and *papa* but also *hija, oh-dear, baby,* and so on).

The articulatory filter provides one interpretation of the mechanism that translates that advantage into a solid advance. And then we see, in the later words, something of the basis for the exponential increase in word comprehension in this period – the widening of phonetic and phonotactic resources, which can in turn support better memory for novel forms, as demonstrated in word learning experiments. And we can also see, in the later words, the point that seems most important: The relative systematicity of these forms reflects the emergence of patterns that are individual by child and that reflect *both* that child's vocal repertoire, as developed in babbling and the first few words, *and* the impact of frequently occurring patterns or word forms in the adult input. Thus while children acquiring the same language remain dissimilar at this early stage, aspects of their phonological patterning can be seen to relate to their common adult language.

References

Baddeley, A. D., Vargha-Khadem, F., and Mishkin, M. (2001) Preserved recognition in a case of developmental amnesia. *Journal of Cognitive Neuroscience, 13*, 3, 357-369.

Ballem, K. D. and Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language, 32,* 159-173.

Barrett, M. (1995). Early lexical development. In P. Fletcher and B. MacWhinney (Eds.), *Handbook of Child Language* (pp. 362-392). Cambridge, MA: Blackwell.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols:.* New York: Wiley.

Boysson-Bardies, B. de, Hallé, P., Sagart, L. and Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*, 1-17.

Boysson-Bardies, B. de and Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*, 297-319.

Braine, M. D. S. (1994). Is nativism sufficient? *Journal of Child Language, 21,* 9-31.

Brent, M. R. and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33-B44.

Cheek, A., Cormier, K., Repp, A., Meier, R. P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. *Language, 77*, 292-323.

Davis, B. L. and MacNeilage, P. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica, 57*, 229-241.

DePaolis, R. A. (2006). The influence of production on the perception of speech. In Bamman, D., Magnitskaia, T. and Zaller, C. (eds.), *Proceedings of the 30th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Deuchar, M. and Vihman, M. M. (2005). A radical approach to early mixed utterances. *International Journal of Bilingualism, 9,* 137-157.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 305-352.

Fennell, C. T. and Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46,* 245-264.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419-439.

Ferguson, C. A., Peizer, D. B., and Weeks, T. A. (1973). Model-and-replica phonological grammar of a child's first words. *Lingua, 31*, 35-65.

Ganger, J. and Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology, 40,* 621-632.

Gómez, R. L. and Gerken, LA. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109-135.

Hallé, P. and Boysson-Bardies, B. de. (1994). Emergence of an early lexicon. *Infant Behavior and Development, 17*, 119-129.

Hallé, P. and Boysson-Bardies, B. de. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development, 19,* 435-451.

Harris, M., Yeeles, C., Chasin, J., and Oakley, Y. (1995). Symmetries and assymmetries in early lexical comprehension and production. *Journal of Child Language, 22*, 1-18.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language.* Cambridge, MA: MIT Press.

Jusczyk, P. W. and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., Cutler, A. and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675-687.

Karmiloff-Smith, A. (1992). *Beyond Modularity* . Cambridge, MA: MIT Press.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). The Head Turn Preference Procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111-116.

Keren-Portnoy, T., Depaolis, R. A., and Vihman, M. M. (2005). The articulatory filter and the creation of sound-meaning links. Presented at the conference on Emergence of Linguistic Abilities, Lyon.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition, 83*, B35-B42.

Knowlton, B. J. and Squire, L. R. (1993). The learning of categories. *Science, 262,* 1747-1749.

Kõrgvee, K. (2001). Lapse sõnavara areng vanuses 1;8-2;1 ('A child's lexical development, 1;8-2;1'). Bachelor's dissertation, Tartu University.

Kuhl, P.K. and Meltzoff, A.N. (1988). Speech as an intermodal object of perception. In A. Yonas (ed.), *Perceptual Development in Infancy. The Minnesota Symposia on Child Psychology, 20.* Hillsdale, NJ: Lawrence Erlbaum.

Locke, J. (1986). Speech perception and the emergent lexicon. In P. Fletcher and M. Garman (eds.), *Language Acquisition* (2d ed.). Cambridge: The University Press.

Macken, M. A. (1979). Developmental reorganization of phonology. *Lingua, 49*, 11-49.

Macken, M. A. (1980). Aspects of the acquisition of stop systems. In G. Yeni-komshian, J. F. Kavanagh and C. A. Ferguson (eds.), *Child Phonology, I: Production*. New York: Academic Press.

Majorano, M., Keren-Portnoy, T. and Vihman, M. M. (in press). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language.*

McClelland, J. L., McNaughton, B. L. and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review, 102,* 419-517.

McCune, L. (2006). Dynamic event words. *First Language, 26,* 233-255.

McCune-Nicolich, L. (1981). The cognitive basis of early relational words. *Journal of Child Language*, *8*, 15-36.

McCune, L. and Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language and Hearing Research, 44,* 670-684.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (ed.), *Language Production*, vol. 2. London: Academic Press.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91,* 3-19.

Murphy, K., McKone, E., and Slee, J. (2003). Dissociations between implicit and explicit memory in children: The role of strategic processing and the knowledge base. *Journal of Experimental Child Psychology, 84,* 124-165.

Nazzi, T., Jusczyk, P. W. and Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43,* 1-19.

Ninio, A. (1992). The relation of children's single word utterances to single word utterances in the input. *Journal of Child Language, 19,* 87-110.

Ninio, A. (1993). Is early speech situational? In D. J. Messer and G. T. Turner (eds.), *Critical Influences on Child Language Acquisition and Development.* New York: St. Martin's Press.

Ninio, A. (2001). Pragmatic keywords and the first combining verbs in children's speech. *First Language, 21*, 433-460.

Nissen, M. and Bullemer, P. (1987). Attentional requriements of learning. *Cognitive Psychology, 19,* 1-32.

Oller, D. K. (2000). *The Emergence of the Speech Capacity.* Mahwah, NJ: Lawrence Erlbaum.

Oller, D. K., Wieman, L. A., Doyle, W. J., Ross, C. (1976). Infant babbling and speech. *Journal of Child Language, 3*, 1-11.

Oviatt, S. (1980). The emerging ability to comprehend language. *Child Development, 50*, 97-106.

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

Priestly, T. M. S. (1977). One idiosyncratic strategy in the aquisition of phonology. *Journal of Child Language, 4*, 45-66.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855-863.

Ruddy, M. G., and Bornstein, M. H. (1982). Cognitive correlates of infant attention and maternal stimulation over the first year of life. *Child Development, 53,* 183-188.

Saffran, J. R., Aslin, R. N. and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A. and Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101-105.

Salo, A. (1993). Muutelõppude ilmumine ühe eesti lapse keelde vanuses 1;5-2;5. [The emergence of inflectional endings in the language of one Estonian child aged 1;5-2;5.] Undergraduate thesis, Finno-Ugric Languages Department, Tartu University.

Schwartz, R. and Leonard, L. (1982). Do children pick and choose? *Journal of Child Language, 9*, 319-336.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195-231.

Squire, L. R. and Kandel, E. R. (1999). *Memory: From mind to molecules.* New York: Henry Holt.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321-1337.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46,* 265-294.

Tomasello, M. (2003). *Constructing a Language.* Boston, MA: Harvard University Press.

Vihman, M. M. (1976). From prespeech to speech: on early phonology. *Stanford Papers and Reports on Child Language Development, 12,* 230-244.

Vihman, M. M. (1981). Phonology and the development of the lexicon. *Journal of Child Language*, *8*, 239-264.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61-82.

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

Vihman, M. M. (1999). The transition to grammar in a bilingual child: Positional patterns, model learning, and relational words. *International Journal of Bilingualism 3*, 267-301.

Vihman, M. M. (2001). A cross-linguistic study of the acquisition of geminate consonants. End of award report, Economic and Social Research Council Award R000223134.

Vihman, M. M. and Boysson-Bardies, B. de. (1994). The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica*, *51*, 159-169.

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

Vihman, M. M. and Kunnari, S. (2006). The sources of phonological knowledge. *Recherches Linguistiques de Vincennes, 35,* 133-164.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., Miller, J. (1985). From babbling to speech. *Language, 61*, 397-445.

Vihman, M. M. and McCune L. (1994). When is a word a word? *Journal of Child Language, 21,* 517-542.

Vihman, M. M. and Nakai, S. (2003). Experimental evidence for an effect of vocal experience on infant speech perception. In M. J. Solé, D. Recasens and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 1017-1020).

Vihman, M. M., Nakai, S., and DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen and C. Best (eds.), *Laboratory Phonology 8* (pp. 341-366). Mouton de Gruyter: New York.

Vihman, M. M., Nakai, S., DePaolis, R. A., and Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language, 50,* 336-353.

Vihman, M. M. and Velleman, S. L. (2000). The construction of a first phonology. *Phonetica, 57,* 255-266.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*, 286-350.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics, 7,* 179-211.

Werker, J., Fennell, C., Corcoran, K. and Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size, *Infancy. 3*, 1-30.

Whalen, D. H., Levitt, A. G. and Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18,* 501-516.

Wijnen, F., Krikhaar, E. and Den Os, E. (1994). The (non)realization of unstressed elements in children's utterances. *Journal of Child Language, 21,* 59-83.

Willingham, D., Nissen, M. and Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 15,* 1047-1060.

Wilson, P. H., Maruff, P. and Lum, J. (2003). Procedural learning in children with developmental coordination disorder. *Human Movement Science, 22,* 515-526.

Figure 1. Model of learning: Procedural (implicit) and declarative (explicit) sources of phonological knowledge

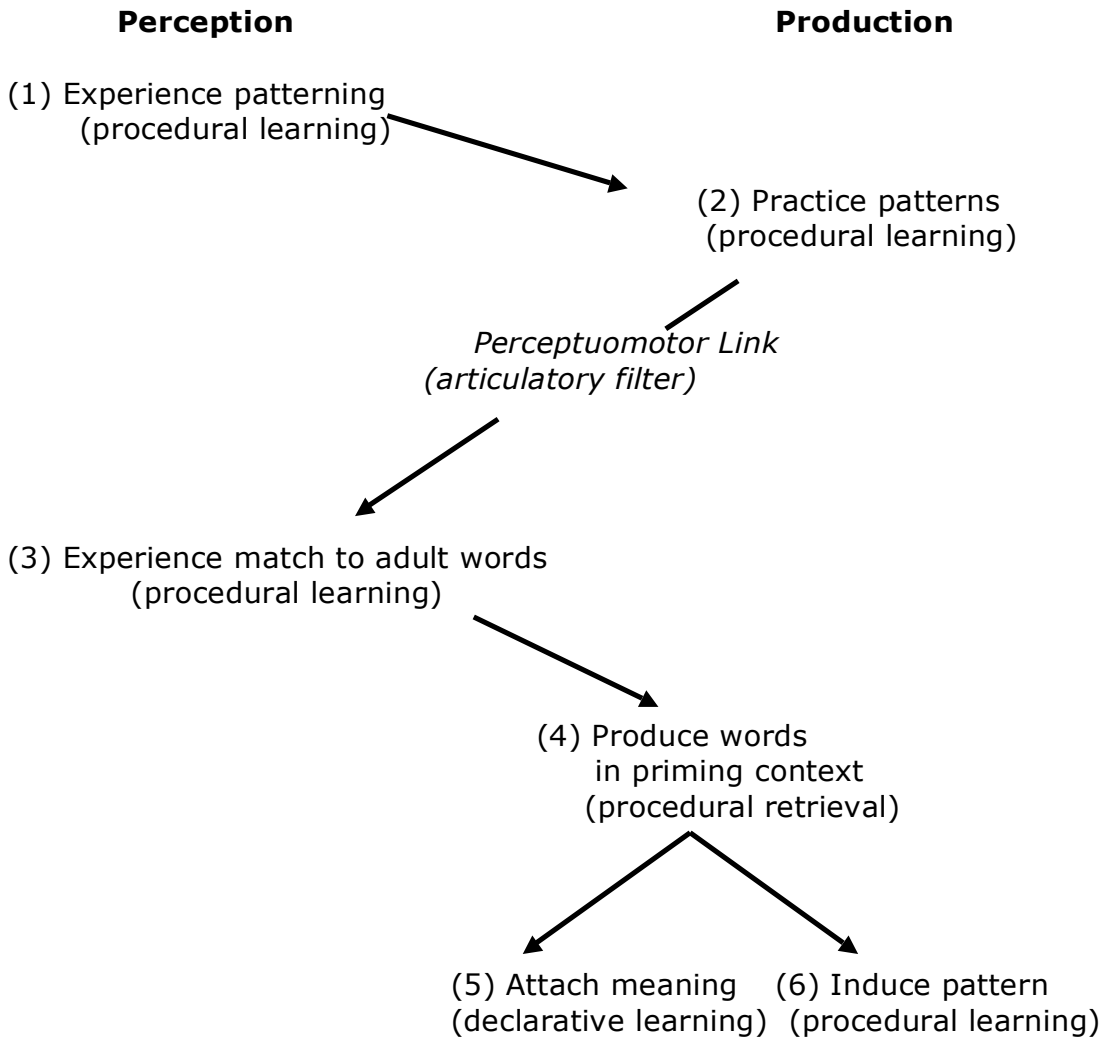

Table 1

Early words

Target words that the child's form closely matches are in bold; words that the child changes by virtue of more radical changes than omission or simple consonant substitution – i.e., words that show reduplication or harmony – are in italics. There are no cases of metathesis in these first words.

a. British English (North Wales; see Keren-Portnoy, DePaolis and Vihman, 2005): first 10 words produced, based on bimonthly 30minute recording sessions.

Jude (10-11 mos.)		**Ian (10-11 mos.)**		**Ali (13-15 mos.)**		**Helena (19-23 mos.)**	
mum	mʌm	**yeah**	jæ	[peek-a-boo, I] **see ya**	diː jaːː, diːːjæʔ, di^{t}da	**mum**	mɑm
hiya	aɪja	who's [th]at?	həʒa, huʒa, husə	good-girl	kʌ^{k}kɑː	**car** [ka:]	ikɑː
cat	ka	bang (bang)	ʌba, babə, papa	**see-saw**	didaʔ, diːdɑː, di^{t}dɑː	**bird**	bəːɪd
[grand]**ma**	ma	catch	kæ, kɛ, kə	gone	gʌ, ga	**I see**	aisiː
barnaby	babi	**hiya**	hʌɪ, heja, heaː	**oh dear**	ʌdɛː	**baby**	beibi
caterpillar	biɔ	**ball**	bo, bo^{l}, bul, bul^{l}, bu^{l}l	**hi(ya)**	aja, ai		
yes, yeah	jə	**up**	ʔʌp				

b. Estonian.

Sources: Vihman, 1976 (Virve); Vihman, 1981 (Raivo); Kõrgvee, 2001 (Madli); Salo, 1993 (Eriku): first words produced, based on diary records. Underlined words are English.

Virve (10-12 mos.)		**Raivo (13-14 mos.)**		**Eriku (from 17 mos.)**		**Madli (11-16 mos.)**	
<u>hi</u>	[hai]	<u>shoe</u>	[ʃ, ç]	tiss ‘nipple’	[tss]	**auto** ‘car’	[auto]
pai ‘nice’	[pai]	viska ‘throw’	[is, iɬ, ɬ]	*päkapikk* ‘elf’	[pæpa]	**kaka** ‘poo’	[kaka]
aitäh /aitæh/ ‘thanks’	[aita, aida]	põmm ‘boom’	[bm bɨm, bɨmː]	tita ‘girl’	[tit]	**kuu** ‘moon’	[kuˑːː]
allo ‘hello (telephone)’	[ao]	aitäh /aitæh/‘thanks’	[ta, taʔ]	*paber* ‘paper’	[paba]	kass ‘cat’	[asˠˡ]
see ‘this’	[se]	**ei** ‘no’	[eiː]	onu ‘uncle’	[enː]	**auh** ‘woof’	[auh]
tere ‘hello’	[te, teðe, tete]	vesi ‘water’	[s̩]	suur ‘big’	[uː]	naba ‘belly button’	[aba]
kikerikii ‘cock-a-doodle-do’	[titiˡ]	pall ‘ball’	[bæ, pæ bæbæ]	väike ‘small’	[æː]	**nämm** ‘yum’	[næmː]
habe ‘beard’	[abə]	<u>hiya</u>	[aja]	vanaema ‘grand-mom’	[ana]	**aiai** ‘ow’	[aiai]
<u>cookie, cracker</u>	[kɔ̰kɔ̰]	banaan ‘banana’	[ba, babæ]	ammuu ‘moo’	[amː]	kott, kotid ‘bag, bags’	[kotˠi]

c. Italian (adapted from Majorano, Keren-Portnoy and Vihman, in press): first 10 words produced, based on bimonthly 30-45 minute recording sessions. Key: m. = masculine, f. = feminine; ipv = imperative.

Anna (10-13 mos.)		**Luca (10-13 mos.)**		**Nicola (10-16 mos.)**		**Nina (13-18 mos.)**	
mamma 'mummy'	mam:a	mamma 'mummy'	mom:œ	**mamma** 'mummy'	mama	**mio** 'mine (m.)'	mio
bebè 'baby	bebɛ	**bella** 'pretty/nice (f.)'	beja	**nonna** 'granny'	nen:a, non:a	baubau 'bowwow'	ba:ˈba:
nonna 'granny'	non:a	**bimba** 'child (f.)'	bimba ə'child f.'	**nanna** 'to sleep (BT)'	nan:a	mamma 'mummy'	mem
nanna 'to sleep (BT)'	nan:a	cocò 'hen (BT)'	kaˈka	**papà** 'daddy'	pap:a	zia 'aunt'	ia
caffè 'coffee'	kakɛ	**pappa** 'food (BT)'	pap:a	bimba 'child (f)'	bœb:ɛ	**bimbo** 'child (f.)'	bib:o
papà 'daddy'	pa:pa	**mimì** 'candy, sweet'	miˈmi	**tata** 'auntie'	tata	caffè 'coffee'	aɛ
		occhio 'eye'	aˈgo	vale 'goodbye'	ae		
		acqua 'water'	akwa				
		tata 'child (BT)'	tit:a				
		bebè 'baby'	beˈbɛ				

Table 2

Later words

Child words that closely matches the target are termed 'select(ed)'; words that the child changes systematically, so that there is more internal similarity among child word forms than between child and adult target, are termed 'adapt(ed)'. Im. = imitated.

a. British English (North Wales; see Keren-Portnoy et al., 2005). The words listed for each child are taken from the first biweekly session with close to 25 words.

Jude (15 mos.) T = 43 words				**Ali (25 mos.) T = 28 words**			
<CV>				**<(CV)CVC>**			
SELECT		**ADAPT**		**SELECT**		**ADAPT**	
blue	[bʌ]	*ball*	[bɔ]	*back*	[pæʲkʰ, baʲk]	*(another) one*	[məm, məməm]
car	[kʰaː]	*book* (im.)	[bʊ]	*bed*	[pat, bad, bət, baʲç, bəbaʲtʰ]	*biscuit*	[əpɪʃ]
no	[nɔ, doː]	*cake*	[kʰi]	*crash*	[oːən kəɬː]	*Tamar*	[nĭmaːt]
sky	[kʰa]	*cheese* (im.)	[kʰi̥]	*feet*	[əhˀmːiːtʰ]	*yeah*	[jˑam]
square	[wɛ]	*flower*	[la]	*mam*	[maˑmː]		
star	[daː]	*mouth*	[ma]	*mess*	[mːæːs, mæːːɬ]		
ta	[tʰaː]	*teeth*	[tiːː]	*pump*	[bɔːp]		
there	[dɛː]						

b. Estonian.

Sources: Vihman, 1996 (Virve), 1981 (Raivo). Diary data are chosen to reflect imposition of a template on adult words (i.e., of 'adaping' targets to child system) within first 50-word period. English words are underlined. F = fricative, N = Nasal

Virve (17-18 mos.)				**Raivo (15-16 mos.)**			
<Ca/ɔ s/n i/u s/n i/u> [see text]				**<C^1VC^1 >**			
SELECT		**ADAPT**		**SELECT**		**ADAPT**	
banaani 'banana'	[paːnini]	*tagasi* 'back (verb prt.)'	[tasisi]	*kiik* (I) 'swing'	[kikː]	*lind* (I) 'bird'	[nɪŋ]
		lennukit 'airplane, obj.'	[nanunu]			*rind* 'breast'	[nən]
		maasikas 'strawberry'	[maːsini]			*king* 'shoe'	[nɪŋ, nɪn, nɛŋ, næŋ]
		porgandit 'carrot, obj.'	[pɔnini]			*banaan* 'banana'	[pamː, bamː, pap]
		raamatut 'book, obj.'	[maːnunu]			*karp* 'small box'	[pap]
		rosinad 'raisins'	[oːsini]			*kits* 'goat'	[tits, titˠ]
						trepp 'step, stair'	[papː]

c. Italian (adapted from Majorano et al., in press): The words listed for each child are taken from the first bimonthly 30-45 minute recording session with 20 or more words. (s1/2/3 'first/second/third person singular'; pres. 'present tense'; f. 'feminine'.)

Anna (20 mos.) T = 27 words				**Nicola (22 mos.) T = 25 words**			
< C_1VC_2V >				**<VCV>**			
SELECT		**ADAPT**		**SELECT**		**ADAPT**	
bimba 'child (f.)'	[bimba]	*ancora* 'again'	[kora]	*alta* 'high'	[atːa]	*cade* 'fall down, s3 pres.'	[ade]
cadi 'fall down, s2 pres.'	[kadi]	*animali* 'animals'	[mali]	*aria* 'air'	[aja]	*cavallo* 'horse'	[alːo]
cane 'dog'	[kane]	*coperta* 'blanket'	[pɛtːa]	*erba* 'grass'	[ɛbːa]	*chiudo* 'close, s1 pres.'	[udu]
coda 'tail'	[koda]	*maialone* 'big pig'	[mone]			*indietro* 'back'	[etro]
dritte 'straight'	[ditːe]	*seduta* 'sit, s3 pres.'	[duta]			*pronto* 'hello (telephone)'	[onto]
gallo 'rooster'	[galːo]					*trattore* 'tractor'	[are]
mucca 'cow'	[mukːa]					*vieni* 'come, s2 pres.'	[ɛni]
metto 'put, s1 pres.'	[metːo]					*zitto* 'silent'	[itːo]

[1] Another characteristic of this kind of learning is its relative inflexibility of access (Squire and Kandel 1999). That is, what has been learned can be retrieved only under closely matching conditions; there is no generalisation to categorically related or similar items.

[2] Following Squire and Kandel I use the terms 'procedural' and 'declarative' here, in lieu of the more widely used terms 'implicit' and 'explicit', to avoid the difficulties of applying the notion of 'explicit' or 'conscious awareness' to infants for whom no such awareness can be clearly established.

[3] The Estonian children were all recorded by their mothers, in the diary study tradition; only these four data sets were available for first words. My two children had exposure to English as well as Estonian, although Estonian was the primary language used in our home. Further illustration of the extent of accuracy in the first words of children acquiring a range of languages can be found in Vihman (1996), Appendix B, which includes the first few words reported for children acquiring Dutch, French, German, Swedish and Japanese in addition to English and Estonian and in Vihman and Kunnari (2006), which reports early words in Finnnish and Welsh in addition to English and French.

[4] 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. Stops were used in the critical VMS contrast to avoid issues of perceptual salience that arise in contrasting a stop and a nasal, for example.

[5] Note that the children range in age from 15 to 25 months; it is unwise to use actual chronological age or estimates as to the timing of the hypothesized 'vocabulary spurt' as the basis for cross-child or crosslinguistic comparisons in this period of rapid change and strong individual differences.

[6] The shift in vowel length from medial to initial syllable is in conformity with the vast majority of common Estonian words, which have initial stress; only stressed syllables have contrast in length.

[7] Only one trisyllabic word – *nii moodi* 'like this' [mi:mona] – produced in the period represented in the table (17-18 mos.) failed to fit the template described here. Of just seven additional words recorded as fitting into a long-word template in the following two months, three have both /s/ and a nasal, but only one of those – *mesilane* 'bee' [mesini] – incorporates the <s…n> pattern while the others harmonise to /s/ or to /n/. One more shows the <a…u> pattern: *magustoit* 'dessert' [masusu], while two (*pikkali* 'lying down' and *vikerkaar* 'rainbow') represent a new departure, but with harmony again affecting the unstressed syllables ([…k:akai] in both cases). See Vihman (1996: 224f.) for these later examples.