

The Acquisition of Phonological Inventories*

Ewan Dunbar and William Idsardi

August 11, 2012

1 The traditional views

The acquisition of phonological inventories is a subject which has been studied by both linguists and psychologists, and rightly so—there is no question from the point of view of the theoretician studying universal grammar that the path from the initial state to the adult state is of interest, and there is no question from the developmental psychologist’s point of view that the child’s phonological capacities are undergoing substantial development in the early years.

Beyond similar titles, however, the linguist and the psychologist researching “phonological inventory development” traditionally have little to share, because by “inventory development” they usually mean quite different things. The tradition among linguists began with Roman Jakobson, who placed the empirical focus on the child’s improving capacities for producing sounds. For psychologists, however, the seminal work of Eimas *et al.* 1971 shifted the focus from observational studies of production to laboratory work in perception. Since then, there have been two different traditions, *child phonology* and *infant speech perception*, both of which use the term “inventory development,” but which have proceeded independently. With two traditions, there come two sets of received facts. The received facts here both come in the form of developmental sequences, and traditional ways of understanding those developmental sequences. We begin by summarizing these traditional views of inventory development from the point of view of the linguist and the psychologist.

1.1 Production: the linguist’s view

A child phonologist asked about the seminal works in the field will likely cite Jakobson’s (1941) *Kindersprache, Aphasie, und Allgemeine Lautgesetze*, first published in English in 1968 as *Child Language, Aphasia, and Phonological Universals* (Jakobson 1968). The *Kindersprache* presented an enticing theory of inventory development that elegantly tied together the three elements of its title. It was so enticing, and had such scope, that it became the standard theory in the study of phonological acquisition,

*This work was partially supported by NSF IGERT DGE-0801465 to the University of Maryland, NIH 7R01DC 005660-07 to William Idsardi and David Poeppel, and SSHRC Doctoral Fellowship 752-2011-0293 to Ewan Dunbar.

gaining a central place in language acquisition textbooks, attracting a search for counterexamples to the empirical claims it contained, and detracting attention from the areas of development it did not cover (see Menn 1980 for a brief summary of this period).

Jakobson's view of acquisition was that phonological inventories were acquired by repeated division of the phonological space into two-way contrasts (see also Drescher 2009). The child would first distinguish vowels from consonants; any further details would be missing until the next contrast was established—for vowels, for example, the next contrast was said to be between low vowels and high-front vowels, and for consonants, the second contrast was said to be between nasal and oral stops. The order of acquisition was claimed by Jakobson to be universal, a claim that was based on a survey of the then-available empirical literature.

At the center of the *Kindersprache* theory was a set of “structural laws” that gave priority to one contrast over another; intuitively, the structural laws gave an abstract complexity measure distinguishing “less-” from “more-structured” sounds. The elegance of the theory was that the same structural laws were said to govern all three of the title areas. The order in which sounds were acquired was said to be the reverse of the order in which sounds were lost in aphasia (“last in, first out”). The structural laws, too, gave rise to a set of crosslinguistic tendencies of the kind that would later come to be called “implicational universals” following Greenberg's (1963) paper: “the opposition of a stop and an affricate in the languages of the world implies the presence of a fricative of the same series,” wrote Jakobson (Jakobson 1968:56)—we conclude that the stop/fricative contrast has priority over the fricative/affricate (or stop/affricate) contrast. It follows from this that fricatives should appear earlier than affricates in language acquisition, and affricates should be lost prior to fricatives in aphasia.

The *Kindersprache* theory was simple and powerful, and, for researchers wishing to pursue it empirically, the theory cut an obvious path, essentially predicting a universal sequence in segmental acquisition. First, was there really a universal order of acquisition? If not, the *Kindersprache* theory as stated was too strong. Second, if there was a universal order of acquisition, what was it? Subsequent questions would depend on having answers to these, and so an empirical literature emerged reporting longitudinal data on the child's changing set of contrasts.

Over the following decades, longitudinal production data, combined with similar data from clinical studies of abnormally developing children, was often pooled in survey papers which attempted to find the commonalities in the observed data. We give an outline compiled from some of these surveys which relate strictly to the development of consonant contrasts (Grunwell 1981, Grunwell 1982, Dinnsen 1992) in Table 1.¹

¹The most comprehensive set of empirical claims about vowel development is still Jakobson's: first, a high/low split ([i] versus [a]), followed by a front/back split (adding [u]), or a secondary height contrast (adding [e]). Jakobson's empirical claims are worthy of serious scrutiny, however, and thorough work describing the emergence of vowel productions is more scarce than work on consonants. The phonetic study of Lieberman 1980 does not distinguish words from babbling, and, more importantly, contains no record of the intended pronunciations in words, meaning that we cannot evaluate the child's *contrastive* inventory; other work is not longitudinal (Davis and Macneilage 1990) or is unfortunately confounded (the subject in Major 1976 is English-Portuguese bilingual). The most usable data for English are to be found in Otomo and Stoel-Gammon 1992; the main finding in that paper, which examines only unrounded English vowels, is that the lax vowels [ɪ] and [ɛ] become contrastive only relatively late, while tense [i] and [a] are contrasted earliest (corroborating Jakobson), with [e] and [æ] falling somewhere in between.

Contrasts	Example Inventory			
Oral/nasal, obstruent/sonorant, coronal/labial ↓	p	t		
	m	n		
	w	j		
Oral/nasal, obstruent/sonorant, coronal/labial, voiced/voiceless ↓	p ^h b	t ^h d		
	m	n		
	w	j		
Oral/nasal, obstruent/sonorant, coronal/labial/ dorsal , voiced/voiceless ↓	p ^h b	t ^h d	k ^h g	
	m	n	ŋ	
	w	j	h	
Oral/nasal, obstruent/sonorant, coronal/labial/dorsal, voiced/voiceless, stop/fricative ↓	p ^h b	t ^h d	k ^h g	
	f v	s z		
	m	n	ŋ	
	w	j	h	
Oral/nasal, obstruent/sonorant, coronal/labial/dorsal/ palatal , voiced/voiceless, stop/fricative ↓	p ^h b	t ^h d	tʃ ʤ	k ^h g
	f v	s z	ʃ ʒ	
	m	n		ŋ
	w	j		h
↓ l > r > θ, ð ...				

Table 1: Sequence of consonant contrasts, drawing on previous reviews and summaries. Example inventories are given for a typical English-speaking child.

Several questions arise from Table 1: first, we should ask how reliable the original sources are, but the authors have nothing to say about this here. Second, we might ask how comparable the studies are. The short answer is that they are fairly inconsistent in their methodology. For example, many generalizations are taken from clinical studies, while others are taken from studies of normally developing children, a bias that simply reflects the greater contact with speech-language pathology practitioners for atypical children. These are not necessarily compatible; see Chapter XXX (this volume [[i.e. Dinnsen paper]]) for discussion of this and related issues. Another important methodological difference between studies is the criterion for adding a contrast to the table. Should we take the contrast to be in place when the sound is produced? When it is used “appropriately”? If so, what does “appropriately” mean? Do children “have a contrast” between [m] and [w] if they only ever use [w] as a substitute for [l]? Do they have a contrast between [w] and [l] if they substitute [w] for [l] in certain environments, but not others? How far should we go in attempting to find these environments, if the substitutions do turn out to be systematic, and we take this to be crucial in determining contrast?

One somewhat radical answer to the question of what it means for a child to have acquired a contrast was suggested by Smith (1973), who constructed a full phonological grammar for the child’s productions, implying that the absence of a particular contrast on the surface could be treated as an epiphenomenon of a neutralization rule, and not the other way around, as a simple reading of Jakobson would suggest. This would imply that the answer to the question of when children “have a contrast” is really “when they produce each segment in all and only the environments that adults do,” so that tables like Table 1, inspired by Jakobson, would really be only approximations to the full set of relevant data. This really represents a different theory (see the discussion of Smith below); we need not move to this extreme, however, to recognize that compiling a table such as Table 1 requires some set of clear criteria. Many of the original sources do not make their criteria explicit, and the rest are generally inconsistent with each other.

Granting that the studies are comparable and reliable, the question of whether the data confirm or disconfirm Jakobson’s claims is of substantial interest. The highlights of Jakobson’s partial order on consonant contrast development are: nasal and oral stops should be distinguished early; labials and dentals should be distinguished later; these should be distinguished from velars and palatals still later; fricatives should be neutralized to stops early, but not the other way around, as noted above; affricates and fricatives should be distinguished after stops and fricatives; [l] and [r] are distinguished late. Table 1 supports the idea that nasal and oral stops, as well as labial and dental stops, are distinguished early, but it is unclear whether there is an ordering between these two developments. Table 1 supports the idea that the velar/non-velar contrast, as well as the palatal/non-palatal contrast, are made later than the labial/dental contrast (except for the glides). Finally, Table 1 also supports the idea that stops precede fricatives, and that the [l]/[r] distinction is late. There are a few other generalizations that emerge about which Jakobson has nothing to say, such as the relative ordering between the acquisition of velars and the acquisition of palatals (Table 1 suggests that velars precede palatals), the place in the order of acquisitions of a voiced/voiceless distinction (Table 1 suggests it is relatively early), and the fact that, in addition to the relatively late emer-

gence of the contrast between the two, neither [l] nor [r] appears at all before some late stage. On the whole, there is nothing in Table 1 that contradicts Jakobson. Inevitably, however, there are published exceptions to even the general progression given in Table 1 (for example, Prather *et al.* (1975) report consistent acquisition of [ŋ] before the velar obstruents, [w] much later than [j], and [r] and [l] before [z]; Olmsted (1971) reports consistent acquisition of [ŋ] only much *later* than the velar obstruents, and also later than the other nasals; and Vihman *et al.* (1986) report a child with [s] and [ʃ] before any velar stops).

Of course, Jakobson did not just claim that these orders could be found in some language; he claimed they were universal, whereas the studies that underpin Table 1 are taken only from English-speaking children. Similar studies of other languages are rare, but, where they exist, they reveal potentially problematic differences for the *Kindersprache* theory of universality. Macken's studies of Spanish acquisition (Macken 1978, Macken 1979) showed that Spanish children make a continuant–non-continuant distinction early, but not a voicing distinction; she concluded that Spanish children do make a voicing distinction at roughly the same time as English children, but initially realize it as a continuancy distinction because of the allophonic status of the fricatives in Spanish (voiced but not voiceless stops are subject to non-contrastive spirantization). Pye *et al.* (1987) report that Quiché-learning children acquire [tʃ] much earlier than [ʃ], and perhaps even earlier than [p], [t], and [k]; they learn [x] long before [s], which seems to be later than [ʃ]; they learn [l] very early, and [p] and [t] appear to be fairly synchronous, whereas [m] is acquired later than [n]. Similar facts can be adduced for Finnish ([d] is late and [r] early; see Itkonen 1977). Some of these facts are in conflict with generalizations in Table 1, while others are also in conflict with generalizations of Jakobson's. There is room for some such conflicts, because the theory does not state that all orderings must be universal; these questions were not satisfactorily resolved before attention shifted to other types of theories, however (see 2.1 below).

Finally, we might ask whether the source of data, child productions, is the only one we might use. It did not have to be the case that empirical research following Jakobson's program focused on only production, although it did. The core idea of the *Kindersprache* theory makes reference to contrast; but the idea of a hierarchy of contrasts underpinning phonological acquisition, loss, and typology is viable whether we are talking about perception, production, or memory. It is only because Jakobson dismissed the idea of studying perceptual development, pointing out that children could easily distinguish (and presumably remember) minimal pairs of words differing crucially by contrasts they could not yet produce, that he took all his evidence from production; as is often the case, this detail of the original author's views shaped the understanding of the theory. Importantly, Jakobson was correct about certain generalities: although exact ages for individual contrast developments in production studies are variable, one thing which is broadly consistent is that the least developed production inventories are typically seen around 1;5, and acquisition of all contrasts can last years. As we shall see below, however, perceptual development for these basic contrasts is largely adult-like by the time native-like productions begin to take shape.

In summary, to the extent that the primary empirical claim of the *Kindersprache*—that there is a universal acquisition sequence—has been assessed, research has revealed that, while there are generalizations to be made, there are exceptions which are worthy

of explanation. In light of this variability, and particularly in light of the cross-linguistic variability that appears to exist in the ordering of contrasts, it would be reasonable to conclude that the simplest version of the Jakobsonian hypothesis is long disconfirmed: whatever the substantive content of the learning mechanism, it does not consist of a simple “checklist.” (See Ingram 1988b and Edwards and Beckman 2008 for evidence that frequency might be able to explain some of the variance.)

In defense of the theory, however, it is worth pointing out that the differences in methodology across studies make it difficult to assess the facts at all. In what is perhaps a more equivocal defense of the theory, we must point out that the inconsistent notion of what the correct notion of it means for a child to “have a contrast” makes it hard to compare generalizations. The fact that Jakobson is ambiguous on this point means that it is difficult in principle to evaluate the theory. Recent work (e.g., Edwards and Beckman 2008) is encouraging in its use of controlled methodology (laboratory-elicited word productions), but the simple percent accuracy measure used there leaves the question of what should qualify as contrast acquisition open; see Ingram 1988a for some suggestions.

1.2 Perception: the psychologist’s view

If the *Kindersprache* theory was a model for elegance of theory in mid twentieth century phonology, the emerging speech perception literature was surely the corresponding model of empirical rigor. When experiments aimed at uncovering the psychoacoustic basis of speech perception revealed that speakers of different languages respond differently to the same sounds on low-level perceptual tasks, it became difficult to dispute the psychological reality of phonological contrast (Abramson and Lisker 1970). If perception was influenced by the linguistic environment in which a person was brought up, then the natural questions were how and when the ambient language came to impress itself upon the perceptual systems.

The paper that broke the empirical ground in infant speech perception was Eimas *et al.* 1971. Armed with the high-amplitude sucking technique of Siqueland and DeLucia (1969), the researchers were able to measure the discrimination abilities of 1- and 4-month-old infants, who, astonishingly, showed the same pattern of discontinuous perception for VOT as English-speaking adults. The insight that measures of infants’ habituation and recovery could be used as measures of discrimination abilities gave Eimas *et al.* 1971 as much methodological cachet as the *Kindersprache* had had theoretical.

It took some time for a clear picture to begin to emerge. Streeter (1976) demonstrated that 2-month-old infants raised in a Kikuyu-speaking environment showed discontinuous perception for stop voicing with an English-like boundary, despite the fact that the Kikuyu VOT contrast is between prevoiced and unaspirated, not between unaspirated and aspirated like English; Lasky *et al.* (1975) reported a similar result using a heart-rate measure for 4- and 5-month-old Spanish-learning infants, although their ambient language also had a non-English-like VOT boundary, suggesting that the Eimas *et al.* discontinuity might not be have been due to influence of the ambient language. (Crucially, adults from language backgrounds with short-lag VOT boundaries do *not* show English-like perceptual boundaries; see Lisker and Abramson 1970.) The

discovery by Kuhl and Miller (1975) that chinchillas also showed an English-like discontinuity for perception of English stops dealt a serious blow to the idea that very young infants showed true categorical (not simply discontinuous) perception, that is, that their discrimination abilities tracked knowledge of linguistic categories.² It was not until Werker and Tees 1984 (which established the use of head-turn procedures, rather than sucking procedures, for infants of suitable age) that a timeline began to be established. The now-famous 10-month developmental milestone, at which infants show a precipitous drop in perceptual sensitivity to non-native contrasts, was not as early as previous results had suggested, but researchers were still surprised at just how early categorical perception was evident (Werker 1989). With a powerful experimental paradigm, and, now, a powerful experimental result, the study of infant speech perception had become, for the initiated, *the* study of phonological inventory development.

In this field of phonological inventory development, the timelines looked different. Having started from the speech perception literature, and not from the *Kindersprache*, the question of a hierarchy of contrasts never arose, and was never systematically investigated, although clear differences from contrast to contrast in the onset of native-language effects were evident from the start. Instead, the next milestone that was sought was the development of a different *type* of knowledge—lexical categories, that is, categories used in long-term memory storage, rather than categories which might only be perceptual.

The question of when children begin to acquire their native-language lexical categories, and how to elicit behavior that reveals lexical and not phonetic categorization, was first asked in the literature by Shvachkin (1948). To encourage young children (between 0;10 and 1;6) to construe strings of speech sounds as “words,” Shvachkin trained them on novel names for objects by presenting the names alongside the objects in play. In this way, the strings would presumably be encoded in the same way as any item in the lexicon. The children were then asked by the experimenter to pick up the object, to use it in play, and so on. After presenting several objects in this way, two of which formed a minimal pair, the experimenter would make a request for one of these two items; performance in retrieving the item was evaluated. From his data, Shvachkin attempted to find a sequence of lexical-receptive contrasts along the lines of Jakobson’s. This literature saw a resurgence in the 1970s, with experiments carried out by Garnica (1973), Eilers and Oller (1976), and Barton (1976). Some tendencies—like an early contrast between sonorants and obstruents and very late acquisition of voicing of stops—did begin to emerge—but the new interest in these experiments was too short-lived for anything clear to be determined. Eilers and Oller (1976) and Barton (1976) also reported improvement with mispronunciations of familiar words versus trained nonce words.

Perhaps the most important point about these results, however, is that there is an enormous discrepancy with respect to perceptual category development. The youngest children tested by Shvachkin and Garnica were 0;10 and 1;5 respectively, and many of these children performed well only on a fairly restricted set of contrasts (obstruent ver-

²The fact that Streeter 1976 also reported Kikuyu 2-month-olds’ sensitivity to a Kikuyu-like boundary not found in English infants of the same age has been cast off as an anomaly in light of later research, (and indeed, earlier research: Lasky et al. did not find a Spanish-adult-like VOT boundary in older Spanish-learning infants), but it was taken to be significant at the time.

sus sonorant, along with some distinctions among sonorants). Later researchers, with knowledge of the relatively early onset of native-like speech perception, would have thus been forgiven for thinking that the word-learning tasks were simply too complicated and the measures too indirect to be informative.

There was some surprise, therefore, when Stager and Werker (1997) reported that 14-month-olds performed poorly in a modern, laboratory version of the word-learning task. In their methodology, the infant was presented with an auditory nonce-word label, paired with a visually presented novel object. In the two-object variant, a habituation phase with two different word-object pairs was followed by a test pair in a ‘same’ condition—one of the previously presented objects was presented with its previous label—or a ‘switch’ condition—one of the previously presented objects was presented with the other previously presented label. A difference in looking times in the two conditions implied successful discrimination.

Interestingly, English-learning 14-month-olds failed in this task when the labels were minimal pairs (/bɪ/ versus /dɪ/). Meanwhile, they succeeded when the words were presented without the objects during habituation (that is, with a checkerboard pattern rather than object pictures, with the transition now being from a one-word habituation phase to a second item at test). Even more interestingly, 8-month-olds performed well in a similar task—a one-object variant—in which only one word-object pair was presented in the habituation phase. (The younger infants could not do the two-object task at all.) Numerous variants on this experiment were subsequently reported, with the original result often corroborated, but not always (Pater *et al.* 1998; Werker *et al.* 1998; Swingley and Aslin 2000; Swingley and Aslin 2002; Werker *et al.* 2002; Fennell and Werker 2003; Fennell 2004; Wales and Hollich 2004; Fennell 2006; Fennell *et al.* 2007; Thiessen 2007; Yoshida *et al.* 2009).

The difficulties did not seem on the surface to be as severe as those of Garnica’s subjects, who struggled to perform on many contrasts up to at least 2;0; in this task, infants seemed to be fully recovered by 1;8, at least on the limited set of contrasts tested (Werker *et al.* 2002; Thiessen 2007). Furthermore, small task differences (including the difference between novel and familiar words explored in the earlier literature) made a big difference in performance. The consensus quickly emerged that the results had something to do with word learning (the “word effect”), but whether there was a true representational failure or some other problem in lexical access or learning was up for debate.

The speech perception paradigm for studying inventory development has now made definitive empirical progress; a standard timeline, with the standard understanding of the results, is given in Table 2. The usual understanding of this literature is that infants begin as “universal listeners,” capable of distinguishing any contrast perceptually, and then learn by “learning to ignore”—but what is the explanation?

The speech perception tradition in phonetic inventory learning research is quite different from the Jakobsonian one, and it has a rather different story to tell. Most importantly, the difference between the slow grind of production development and the quick transition from newborn to native-like hearer is empirically undeniable. This simple difference seems to undermine the very premise of the acquisition of *the* phonological inventory.

Age	Perceptual Changes	Contrasts tested
0;1–0;2	Sensitivity to non-native contrasts	English, Kikuyu infants show a categorical VOT boundary for labials at around 25–30 ms, despite the absence of such a boundary in Kikuyu
0;6	Adult-like warping of vowel perception, as revealed by different directional asymmetries in discrimination Continued sensitivity to non-native consonant contrasts, as revealed by discrimination performance similar to adults or older infants from language environments in which the contrasts are native	[i]–[y], for English versus Swedish infants [t ^h]–[t], [kʰ]–[qʰ], [k]–[kʰ], [b]–[b], [ʔ]–[ʕ] in English infants; [r]–[l] in Japanese infants
0;6–0;8	Poor sensitivity to certain difficult native-language contrasts	[f]–[θ], for English-learning infants; syllable-initial [n]–[ŋ], for infants learning Filipino (Tagalog)
0;8	Ability to detect changes in novel native-language minimal pairs associated with objects	[bi]–[di], for English-learning infants
0;10–0;12	Decline in sensitivity to non-native consonant contrasts to adult-like or near-adult-like levels	Contrasts listed at 0;6
0;10–1;2	Improvement in sensitivity to difficult native-language contrasts	Contrasts listed at 0;6–0;8
1;2–1;3	Decline in ability to distinguish novel minimal pairs associated with objects	[bi]–[di]; [bɪm]–[dɪm]; [bɪn]–[p ^h ɪn]; [dɪn]–[p ^h ɪn]; [dɪn]–[gɪn]; [dɑ]–[t ^h ɑ], all for English-learning infants
1;5–1;8	Improvement in ability to distinguish minimal pairs when associated with objects	[bi]–[di]; [dɑ]–[t ^h ɑ], for English-learning infants

Table 2: Milestones in infant speech perception.

2 Revisiting tradition

What does it mean to acquire a phonological inventory? “At first glance,” writes Jusczyk (1992), “. . . [i]t would seem to be a matter of identifying the elementary sound units that are used to form words in the language.” This statement might seem innocuous and neutral. In fact, however, it represents only one answer to the question of what it means to acquire a phonological inventory. One can easily identify at least three others.

We find it helpful to think of human speech-sound cognition (*phonology*, in a broad sense) in terms of three inescapable facts: humans have ears; humans have mouths; and humans have memories. Phonological processing includes, minimally, some receptive processing system, some productive processing system, and some storage system. Processing information entails having some way of encoding that information. Minimally, then, there are three encoding formats used by the brain to process speech; these three formats are *conceptually* distinct, even if they are not all *actually* distinct. Add to this the observation that the storage system must interface with both the productive and the receptive systems, and we obtain four different possible senses of the term “phonological inventory”: a set of possible distinctions that can be made when perceiving speech (the *phonetic–receptive* inventory); a set of distinctions that can be made when storing perceived speech in the lexicon (the *lexical–receptive* inventory); a set of distinctions that can be made when producing speech (the *phonetic–productive* inventory); and a set of distinctions among speech sounds that can be made when storing instructions for producing words in the lexicon (the *lexical–productive* inventory). These four senses of “phonological inventory” are summarized in Table 3.

LEXICAL–RECEPTIVE: Memory encoding (perceived speech)	LEXICAL–PRODUCTIVE: Memory encoding (speech to be produced)
PHONETIC–RECEPTIVE: Information used in speech processing	PHONETIC–PRODUCTIVE: Information used in speech production

Table 3: Four conceptually distinct representations that have been called “phonological inventories” in the developmental literature.

For example, linguists will be familiar with the lexical versus phonetic distinction made in Table 3 (vertical). Grammatical descriptions constructed by linguists are usually viewed as descriptions of a mapping between these two cognitively distinct types of representations, and it is for this reason that finding an adequate theory of these grammars is a part of cognitive science (see Chomsky and Halle 1968 and Prince and Smolensky 2004). On the other hand, most theoretical works in phonology assume a

single set of features common to perception and production (likely because of the influence of Jakobson *et al.* 1952) implying that the receptive versus productive dimension in Table 3 (horizontal) is not relevant in the lexicon. Psychologists, however, are likely to be familiar with theories that the lexicon contains two separate sub-stores, one for recognizing language, and one (linked, but in some sense distinct) for producing language (Straight 1980; Caramazza 1988; see Menn 1992 for such a proposal approached from within theoretical linguistics), which would require two separate encodings at the lexical level, and thus in principle two separate developmental tracks. The phonetic levels are simply those representational formats that the brain uses to encode information about perceived speech and to control motor systems. A multitude of logically possible distinctions other than these come to mind, but these four senses are the ones most frequently found in the literature and in current thinking on the development of phonological representations.³

Having reviewed some of the classic literature on inventory development, we encourage the reader to reconsider some of the principal results and claims in terms of this four-way distinction. As was made clear above, for example, the child phonology literature has focused on productive representations, while the infant speech perception literature has focused on receptive representations; we might wonder how we can relate the two. More difficult questions arise when we ask whether particular theories or pieces of evidence relate to lexical or phonetic levels, and, although some attempt has been made to dissociate the two in the speech perception literature, we will find reasons to doubt how well we have dissociated them up to this point. In this section, we consider a few of the issues raised.

³All researchers will have heard the term *phoneme* used in one or more of these contexts, but we have avoided it. In certain schools of early twentieth century phonology in which many researchers considered their enterprise to be only loosely connected to cognition (notably, American Structuralism), there were two senses of the term, corresponding to the minimal units on two different levels of analysis (the *taxonomic phonemic* level and the *abstract phonemic* level). Early generative linguistics (e.g., Chomsky 1964; Chomsky and Halle 1968) avoided the term in favor of a distinction between “lexical” (or “underlying”) and “phonetic” representations, except when referring to the work of others. The term “phoneme” quickly worked its way into work in generative phonology, however. These early generative uses of “phoneme” seem to refer to the unit of lexical representation; but, in more recent work, it is sometimes difficult to tell whether the term is defined by the level of representation it refers to or by certain properties that have become associated with that level of representation, and, if so, which properties (for example, division into segments, degree of abstractness, or assumption of a finite inventory). Outside linguistics, the situation is more confusing, as the term is often used to refer to any or all of the four representations in Table 3. For these reasons, we avoid the term to prevent misunderstanding. We also stress that the term “inventory” in this chapter could usually be replaced with “representational capacity”; most of the facts and theories discussed could be restated without assuming a finite inventory of segments (for example, the same themes would be stressed if representational development were considered from the point of view of exemplar theory); nor do we assume that what is developing in the infant are systems for representing individual sounds, like [s] or [i], since what we have presented is too general to bear on the nature of the features that represent individual sounds. We leave it to the reader to consider these issues. Although there is some controversy over these fundamental assumptions, understanding these debates is not crucial to considering most of the developmental literature, and attempting to be entirely neutral would have led to confusing terminological awkwardness. For some discussion of these issues, see Dunbar and Idsardi 2010.

2.1 What level is this? Jakobson and Smith

Current evidence suggests that the child's perceptual abilities are relatively adult-like during the period of interest to Jakobson, in which production seems to be still in flux. Jakobson's discussion and his examples from the observational literature of the time (for example, the one-year-old son of Serbian linguist Milivoj Pavlović (1920), who understood the difference between the words for his father, *tata*, and his excreta, *kaka*, but called them both *tata*), are meant to suggest that both the phonetic–receptive inventory and the lexical–receptive inventory are in place. What inventory was it that Jakobson was watching develop? Following our four-way distinction, it could be the lexical–productive inventory, or of the phonetic–productive inventory, or both, or a mapping between them. In any case, Jakobson's inventory was a *productive* inventory, meaning that, for Jakobson, the productive and receptive inventories were necessarily distinct on some level. Furthermore, according to Jakobson, the productive inventory is not limited just by the child's motor skills, since, in babbling, the child can produce a wide range of native and non-native speech sounds (see also Hale and Reiss 2008). Empirically, this is an overstatement: while children do seem to have (somewhat) larger babbling repertoires than they show in word production, the set of babbling sounds is fairly similar to the set of sounds they use in words (Vihman *et al.* 1986); nevertheless, there is a protracted arc of productive inventory development after babbling ceases which needs to be explained regardless of how and whether it is related to babbling, and Jakobson does not tell us enough to know exactly which parts of the cognitive system this development should be attributed to.

An answer to the question of just what was developing, if not motor skills, arose in the generative tradition shortly after the English publication of the *Kindersprache*. Stampe (1969) and Smith (1973) replaced Jakobson's theory of representational development with a theory of grammatical development. They attempted to characterize a level of grammatical preprocessing prior to the articulatory level, that would map from adult-like stored forms, assumed to be fairly accurate, to child-like productions. They claimed that this computation was the same type of computation that went on in adult phonology (at the time, the kind of computation laid out Chomsky and Halle's (1968) *Sound Pattern of English*).

Both Stampe and Smith presented further arguments against the motor-failure hypothesis. Particularly compelling was Smith's so-called *puzzle-puzzle*. Smith's son, Amahl, at some stage, mapped all adult /d/ to [g] in a certain context (before syllabic [l]). Under a motor-failure theory, this would be because Amahl was unable to coordinate his muscles to pronounce /d/ in this context. But, in fact, Amahl mapped adult /z/ in this context to [d], giving the chain-shift outcome in (2.1):

$$(2.1) \begin{array}{l} /p\Lambda d_l/ \longrightarrow [p\Lambda g_l] \\ /p\Lambda z_l/ \longrightarrow [p\Lambda d_l] \end{array}$$

Smith reports thoroughly testing that Amahl was in fact able to perceive this difference and map it on to different lexical items (though Macken (1980) objects to the claim that Amahl's perception of the contrast was adultlike); his data also demonstrate that these substitutions were systematic. Two similar cases are presented by Smith in

his 1973 book, and other cases have been noted before and since (Aleksandrov 1883; Smith 2010).

The new insight here was that the absence of a contrast between, say, [d] and [g], did not entail that child would simply always select one, or choose one at random. If the child was systematic (therefore, consistent), the limitation could not be in the ability to produce a particular segment, or even in the ability to produce it consistently, and, as the *puzzle*-puzzle demonstrated, the limitation could not be in the ability to produce a particular sequence, either. (Smith also had a way of ruling out the possibility that productive development was due to changing lexical inventories, which would have manifested itself in the form of mislearned lexical items. For details, see Smith 1973; for a response, see Menn 1992; for review of the issue, see Smith 2010.)

Though the details of Smith's analysis were criticized (Braine 1976; Macken 1980; Menn 1980), and though Smith was not the first to attempt a similar analysis (see, for example, Chao 1951), the idea that child productions were the product of systematic substitution rules similar to adult phonological rules, and that misperception was not the source of the majority of children's phonological errors, stimulated a large amount of research. Furthermore, since Smith made his full longitudinal lexicon available in an appendix, many subsequent papers have reanalyzed the Amahl data (Macken 1980; Goad 1997; Dinnsen *et al.* 2001; Vanderweide 2006).

One consequence of Smith's conclusions is that, regardless of whether the lexical-productive inventory is distinct from the lexical-receptive inventory, the phonetic-productive inventory is not the same as the lexical-productive inventory. Under this theory, a Jakobson-style expansion in phonetic-productive representational capacity either drives, or is an epiphenomenon of, the development of the production grammar. Clearly, if the phonetic-productive inventory is epiphenomenal, then it is not an object of study by itself; recently, however, some theories have begun to treat phonetic-productive inventory development as a representational expansion again. The formal treatment has been in terms of a changing set of available features, either with restrictions stated as Optimality-Theoretic (OT) markedness constraints (Boersma and Levelt 2003; Kager *et al.* 2004), or stated directly in a more Jakobson-like theory specialized for markedness relations among features (Rice and Avery 1995; Dresner 2009; related child language analyses are to be found in Levelt 1989; Fikkert 1994; Dinnsen 1996; Brown and Matthews 1997).

Another of these representational approaches (Vihman and Croft 2007; Altwater-Mackensen and Fikkert 2010) locates Jakobson's developmental arc in the lexical-productive inventory, and, furthermore, claims that there is only one type of lexical encoding, shared by the lexical-receptive and lexical-productive functions. The obvious challenge for this approach is to explain why children are capable of discriminating contrasts they cannot produce. One response has been to claim that, unlike children's simple phonetic discrimination performance, children's performance in word-learning tasks does in fact mirror production; see section 2.2 for further discussion.

Finally, it is worth remembering that our four-way division of logically possible inventories is not complete, nor can it ever be; one can always think of finer sub-divisions of each cell, and theories of phonology and speech processing that would make additional nuances critical. Although Smith assumed the child's phonological grammar to be a sequence of rules manipulating discrete feature values, there is evidence

for learned “phonetic processes” operating on sub-symbolic information (Sledd 1966; Dyck 1995; Flemming 2001), in addition to learned cross-linguistic differences in the phonetic implementation of individual phonological features (Pierrehumbert 2003). Taking into account either of these two parts of phonological processing would suggest that theoreticians might consider distinguishing between two types of phonetic-productive inventories, one which is the output of a Smith-type grammar operating over discrete feature values, and one which is the output of subsequent phonetic implementation and phonetic adjustment processes. Since some studies have reported sub-phonemic changes in children’s productions over time (Macken and Barton 1980; Scobbie *et al.* 2000), there are very likely many theoretically important facts waiting to be discovered in the phonetics of child productions.

2.2 Word learning tasks

The standard interpretation of the result of Stager and Werker (1997) discussed in section 1.2 is that infants can be good at perceiving contrasts without being good at storing them in or retrieving them from the lexicon. If we accept these tasks probe something lexical, then we need to ask how it could be that the storage step seems to fail when discrimination is at ceiling.

A familiar explanation is that there are two kinds of encoding under development: both the phonetic-receptive encoding and the lexical-receptive encoding are in flux during infancy, changing in response to the linguistic environment. The explanation here is straightforward: discrimination tasks show the development of the phonetic-receptive encoding (thought of as a set of phonetic categories, distinct from the child’s phonological categories); word-learning tasks show the development of a second encoding, to which phonetic information must be mapped in the lexicon. A variant of this view maintains that there is really only one linguistic level that develops in response to the environment—the lexical-receptive level—and early changes in discrimination performance reflect low-level adaptation of the auditory system. Researchers may decide for themselves whether they believe that “phonetic” and “auditory” mean the same thing in receptive processing, but the choice does not change the general shape of the explanation: if and only if children fail in a word-learning task, their lexical-receptive inventory must be insufficiently developed to represent the contrast being tested.

Taking the novel word-learning data as the primary source of evidence about lexical inventory development implies that the lexical-receptive inventory is acquired relatively late; in particular, although infants apparently fail at Stager and Werker word-learning tasks at 14 months, their first words typically come around 12 months. How can early lexical knowledge exist without an encoding scheme for storing words in the lexicon? A resolution is to be found in a theory that claims that phonemic representations develop in response to an enlarging lexicon (Brown 1973; Charles-Luce and Luce 1990). The claim is essentially that the infant brain is equipped by default with a system which can encode a few words, but which is insufficient to encode a full human lexicon.

All of these views restrict the interpretation of word-learning results to the lexical-receptive inventory. An attempt to relate the word-learning results with the development of the productive inventory has been put forward by proponents of *phonological*

underspecification. This is the view that the encoding format for speech sounds to some degree follows the Saussurean “nothing but differences” principle, encoding certain speech sounds by systematically leaving the values of certain encoding features (dimensions) effectively specified with *none* of the legal contentful values (Lahiri and Reetz 2002; Dresher 2009). For example, it is commonly held that place of articulation can be specified as CORONAL (using the tip or blade of the tongue), LABIAL (using the lips), or DORSAL (using the back of the tongue), but underspecification theories often contend that the feature CORONAL acts as a universal default, and only LABIAL and DORSAL are specified explicitly. Predictions about misperception asymmetries have been derived from these types of claims in the psycholinguistic and neurolinguistic literature, for example, that a larger mismatch negativity (MMN, a neurophysiological change detection response; Näätänen *et al.* 1978) should be observed when a repeated sound with a marked feature value is changed to another sound, as compared to when a sound is changed from one with an unmarked feature value to another sound: because there is no feature to restrict the listener’s expectations about the following sound, a mismatch should not be detected (Friedrich *et al.* 2008; Scharinger *et al.* 2012).

Underspecification of a unified lexical–receptive–productive inventory has been proposed as an alternate explanation for some of the behavior seen in word-learning tasks (Fikkert 2005; Fikkert and Levelt 2008). For example, Altvater-Mackensen and Fikkert (2010) propose that children go through a stage in which all consonants in a word must share the place feature of the vowel (both alveolar consonants and high front vowels are CORONAL, both velar consonants and low vowels are DORSAL, and so on, under the feature theory of Halle *et al.* 2000); if children are in this stage, they may also fail to detect certain minimal pair contrasts, but not others. In particular, a change from /ba/ to /da/ would be detected, because the “correct” feature value for the consonant would be DORSAL (the harmonic form would be /ga/); a change from /bi/ to /di/ would not be detected, however, because the difference between the marked LABIAL feature and the unmarked harmonic CORONAL feature would be undetectable.

Surveying the previous literature, it is certainly true that researchers have not taken into account the place features of the vowel when constructing materials. A survey of all known published studies following the Stager and Werker methodology reveals only the following crucial pairs: /bi/–/di/ (Stager and Werker 1997; Werker *et al.* 2002), /ɪf/–/nim/ (Stager and Werker 1997; Werker *et al.* 1998), /bal/–/dal/ (Fennell and Werker 2003; Fennell 2004), /dal/–/gal/ (Fennell and Werker 2004; Fennell 2004), /dm/–/gm/ (Fennell 2004), /bm/–/dm/ (Pater *et al.* 2004; Fennell 2006; Fennell *et al.* 2007), /bm/–/p^hm/ (Pater *et al.* 2004) /dm/–/p^hm/ (Pater *et al.* 2004) and /t^hɔ/–/dɔ/ (Thiessen 2007). Interestingly, although it is certainly not the case that all of the place contrasts tested with the coronal/front vowels /i/ and /ɪ/ have always failed to be detected using the Stager and Werker methodology, it is notable that both /bal/–/dal/ and /dal/–/gal/ are apparently detectable by 14-month-olds, suggesting a previously overlooked confound between vowel place and word familiarity, the factor to which previous authors attributed infants’ success on this task. Nevertheless, there remain numerous other experimental factors which can give rise to success on many pairs with coronal vowels in the Stager and Werker paradigm, and at least one paradigm (the visual choice paradigm of Swingley and Aslin 2000) in which 14-month-olds seem to perform well overall (Yoshida *et al.* 2009 tested infants in a word-learning task; Swingley and Aslin 2002

tested a wide variety of contrast, but used known words).

Finally, there is yet another way of looking at the data, which denies the tight link between word-learning performance and lexical encoding development. Storing and retrieving lexical items clearly involves more than simply encoding a set of sounds; for one, it involves encoding a *sequence* of sounds, and, for another, it involves exploring the links to the corresponding semantic and morphosyntactic entries (for relevant psycholinguistic theories, see Bock and Levelt 1994; Caramazza 1988). Failure in the broader lexical encoding step, not a deficit in the ability to encode individual segments, was proposed as an explanation for infants' poor performance by Stager and Werker (1997). We are not aware of any experimental study attempting to disentangle these two factors.

This view might help to resolve some empirical puzzles in the word-learning data. For example, there is a difference in performance in word-object pairing tasks between known and unknown words, with performance on known-word/non-word minimal pairs much better than performance on nonce-word minimal pairs (Barton 1976; Swingley and Aslin 2002). If the infant has a perfect ability to store words, up to the limits of her lexical representational capacity, then, once learned, novel words should be the same as known words; the facts could be predicted by a theory positing a deficit in the ability to store words, without any faults in lexical representational capacity. Correctly constructed, such a theory might also account for the fact that seemingly small changes to laboratory word-learning tasks can affect performance greatly, so that performance is not always as bad as might be expected if there were an outright failure to encode particular contrasts (Fennell 2004; Thiessen 2007; see Werker and Fennell 2004 for discussion).

3 The future

To some extent, the goals in the study of phonological inventory acquisition overlap with the goals of the speech sciences as a whole: how is speech represented in memory? For production? For perception? Are some of these representations really the same? If not, how do they interact? Are they categorical or not? Absent answers to these questions, we can only hope that child data will be informative in roughly the same way adult data are. For example, children can recruit surprisingly fine phonetic detail in speech processing (McMurray and Aslin 2005); but the presence of phonetic detail does not entail the absence of coarser encodings in speech processing, or even imply primacy of more detailed encodings. We would also like to see experiments that attempt to determine under what circumstances infants do *not* pay attention to phonetic detail, and, ideally, in what types of representations. (For adults, these have sometimes taken the form of priming studies along the lines of Pallier *et al.* 1999.)

Even issues which seem to be strictly developmental are at heart issues about language processing more generally. When do infants construct higher-order abstractions of speech sounds, and how? Are the word-learning results relevant? This gets at a more fundamental issue—where is the abstraction in speech processing? Is it mainly in the lexicon, with phonetic representations full of detail, or are coarse representations formed early in receptive processing? Similarly, when we ask about the relation

between perception and production in infancy, and the mystery of children’s defective pronunciations, we touch on more basic questions: are lexical representations stated in a receptive alphabet, a productive alphabet, or both, or neither?

In answering these questions, there is only so far the current type of empirical and theoretical literature can go. Although timelines are interesting, the real question in language development is *how* it takes place. In recent years, research on language learnability—the traditional term in linguistics for the theoretical study of learning mechanisms for language—has made inroads into previously uncharted territory by turning to well-understood principles and tools from statistics, machine learning, and the areas of computer science and mathematics related to optimization and search—related fields which could perhaps be collectively referred to as the *inference sciences*. During the 1990s, most learnability research presented algorithms and principles which were built to order for language acquisition problems, and rarely drew explicit connections to these other fields (Dresher and Kaye 1990; Gibson and Wexler 1994; Boersma 1997; Tesar and Smolensky 1998). However, deeper analysis of these and related algorithms (Niyogi and Berwick 1996; Pater 2008; Boersma and Pater 2008; Magri 2012), as well as new applications of standard techniques from the inference sciences (Yang 2002; Goldwater and Johnson 2003; Hayes and Wilson 2008), have helped to underscore the close connection between linguistics and these other fields.

The study of receptive inventory acquisition has been greatly advanced by the simple observation that, at least for receptive inventories, the learner’s problem is one of *clustering* the auditory input. Clustering is a standard problem in machine learning in which, presented with a collection of tokens, the learner must sort out the tokens into some number of categories (Hastie *et al.* 2009); the harder (and unfortunately more realistic) version of this problem also requires determining how many categories to posit. A *mixture model* is the statistical term for the generative model obtained by clustering, in which each token in the input is assumed to be an instantiation of one of a discrete set of categories.

Clustering is different from *classification*, in which the problem is, given a description of some set of categories, to predict the categories of new points, but the two problems are intimately related: the solution to a clustering problem forms the input to a classification problem. Therefore, armed with some classification behavior (say, phoneme classification in adults or infants), plus some hypothesis about what method of classification is being used, we can try to work backwards to determine what techniques for clustering might have been used to arrive at this classification, or what types of information might be useful.

Research in cognitive modeling of phoneme acquisition is in its infancy, however, and has up to now taken a less ambitious approach, simply attempting to find statistical methods that find any sort of phoneme-like categories in acoustic data at all, rather than attempting to match the fine details of phoneme classification. As such, research in the field has often taken the *ideal learner* approach. This means ignoring many of the real-life constraints on the learning algorithm (like memory and speed) and attempting to determine how well a learner could do without these constraints (reminiscent of Chomsky’s (1965) “instantaneous acquisition”). In practice, it means applying standard clustering techniques which are known to find some good or optimal clustering solution and experimenting with different model assumptions and different inputs; if

an ideal learner appears to find categories like those posited by a “gold standard” linguistic analysis under a certain set of modeling assumptions, then a real learner would presumably also benefit from taking the same approach.

Assuming that we can determine the correct statement of the learner’s input at the level of auditory cortex, the clustering problem for receptive inventories is fairly well specified, given the last half century of productive research into the perceptually relevant acoustic dimensions for speech. For vowels, for example, the first through the third formants plus the duration can today be measured readily and fed into one of any number of off-the-shelf clustering algorithms. This simple approach is that taken by DeBoer and Kuhl (2003; for three English vowels, using Expectation Maximization to fit a mixture of Gaussians), by Currie-Hall and Smith (2006; for Greek vowels, using *k*-means clustering), by Vallabha et al. (2007; for Japanese vowels, using both an incremental version of expectation maximization for a mixture of Gaussians and a non-parametric extension of the same algorithm), and by McMurray et al. (2009; for English VOT, using essentially the same parametric mixture estimation algorithm as Vallabha et al. 2007).

Even for vowels, where the relevant acoustic parameters are thought to be well understood, the clustering problem is in general very hard using raw data for current models for systems with more than a few categories. One promising change to the assumptions of the model has been to construct categories which partial out the effects of allophonic rules (or, more generally, any effects of context or other variables), thereby removing some of the noise from the input (Dillon *et al.* 2012). Another potentially promising approach is to add an extra layer to the model corresponding to a set of known words (that is, a lexicon), thereby using context in another way, to help recover misclassified acoustic material by attracting each token (now a word) to a known lexical item (Feldman *et al.* 2009). Both these approaches assume that the encoding of interest is a lexical inventory, suggesting that, from a learnability perspective, it may be unnecessary and even counter-productive to attempt to discover a set of phonetic–receptive categories, rather than simply discovering a set of lexical–receptive categories directly.

Of course, the view that there are two phonetic receptive inventories is the one we would obtain if we directly translated the standard tools for phonemic analysis taught to linguistics undergraduates into a learning mechanism. The analyst must first determine what the possible segments of the language are, including all positional variants (phone discovery); the analyst then discovers phonemes by grouping phones in some way. This might be done agglomeratively, by collapsing certain predictable distinctions—for example, by looking for complementary distributions between pairs of phones (Harris 1951, Peperkamp *et al.* 2006)—or it might be done divisively, by searching for evidence (for example, minimal pair evidence) that a pair of phones is contrastive (Dresher 2009). Grouping phonetic categories in this way is difficult to do well given current approaches to discovering phones (Dillon *et al.* 2012).

A further insight gained by taking a computational perspective is that the problem of inventory discovery is inherently a statistical one—one that requires reasoning under uncertainty—in two senses: first, the discovery procedure must sort out signal from noise, and thus must allow for some uncertainty about the correctness of its solution or the relevance of an individual data point; second, the resulting inventories, at least receptive inventories, seen as classifiers, clearly have regions of uncertainty, a com-

monplace result in adult vowel identification tasks. Most of the techniques which have been applied in this domain (with the exception of k -means and k -nearest neighbors) are statistical in both these respects.

A quickly growing trend in statistical modeling is the application of the Bayesian perspective, which treats probability as a quantification of uncertainty of knowledge rather than the limit of relative frequency students are typically introduced to (see Knill and Richards 1996, Doya *et al.* 2007 for discussion of Bayesian models in perception and brain science; see Chapter XXX in this volume [[i.e. Pearl and Goldwater]] for more detailed discussion of the relevance to acquisition).

Taking a Bayesian perspective makes it licit to state probabilities not only of observable events, but also of hypotheses, and thus to use probability theory as a very generic method for changing information states—that is, learning. The crucial relation in this context is Bayes' Rule, given in (3.1), which states that the probability of a hypothesis H given some set of data D (the posterior probability of H) is proportional to the probability of the data given the hypothesis (the likelihood of the data) times the probability of the hypothesis before seeing the data (the prior).

$$(3.1) \quad \Pr(H|D) \propto \Pr(D|H)\Pr(H)$$

There are some simple but powerful consequences that can be drawn about learning if humans are Bayesian learners. For one thing, each hypothesis is in principle associated with a different likelihood function, and many likelihood functions will assign lower probability to the observed data simply because they are more general: since probabilities must integrate to one, if a hypothesis assigns a large amount of probability to many unobserved events, there is less probability remaining for the observed events. This effect drives a Bayesian learner to be conservative, and the strength of the effect increases as a function of the number of data points (Tenenbaum 1999). However, a Bayesian treatment of learning can also be hierarchical, with complex learning problems depending on quantities which themselves must also be learned; in this case, the same type of principle applies to the intermediate level of the hierarchy, giving an automatic Occam's Razor-like bias toward simpler intermediate-level hypotheses. Learning the parameters of a mixture model while learning the number of categories is one such problem. The likelihood term favors solutions with more categories, because such solutions can provide a more fine-grained (thus, conservative) description of the data; however, most sensible priors will imply for a similar reason a penalty on more complex solutions (that is, with more categories) just by virtue of the fact that they provide more ways of describing the same data. For details of the general point, see MacKay 2003.

Beyond this interesting theoretical point, however, Bayesian modeling is highly practical, in that it allows researchers, in principle, to explore the consequences of learning extremely complex and nuanced models in a relatively "plug and play," non ad-hoc manner (see Griffiths *et al.* 2008 for a review). Given the large number of unresolved theoretical questions discussed, explicitly or implicitly, above, Bayesian models will surely find a prominent place in future inventory acquisition literature. One potential application is in reconciling perception and production: systems in which uncertainty can be quantified seem particularly appropriate for exploring analysis-by-

synthesis type models of speech perception, in which speech is perceived by determining the most likely gesture to have generated it. There are countless other informative projects waiting to be carried out.

We cannot always reach a level of explicitness in our theories which makes them implementable by machine, although there will be a day when this is much easier than it is today. In the meantime, we hope only that acquisition researchers will keep in mind the inescapable facts—that the children under discussion all have ears, mouths, and memories—and will state their objects of study clearly.

References

- ABRAMSON, ARTHUR S., and LEIGH LISKER. 1970. Discriminability Along the Voicing Continuum: Cross-Language Tests. In *Proceedings of Sixth International Conference of Phonetic Sciences*, 569–573.
- ALEKSANDROV, ALEKSANDR. 1883. Detskaja reč'. *Russkij Filosofskij Vestnik* X.86–120.
- ALTVATER-MACKENSEN, NICOLE, and PAULA FIKKERT. 2010. The acquisition of the stop-fricative contrast in perception and production. *Lingua* 120.1898–1909.
- BARTON, DAVID, 1976. *The Role of Perception in the Acquisition of Phonology*. University College, London dissertation.
- BOCK, KATHRYN, and WILLEM LEVELT. 1994. Language production: grammatical encoding. In *Handbook of Psycholinguistics*, ed. by Morton A. Gernsbacher, volume 3, 945–983. San Diego: Academic Press.
- BOERSMA, PAUL. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21.43–58.
- , and CLARA LEVELT. 2003. Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3.1–50.
- , and JOE PATER, 2008. Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar.
- BRAINE, MARTIN D. S. 1976. Review of Neil Smith (1973). The acquisition of phonology: a case study. *Language* 52.489–498.
- BROWN, CYNTHIA, and JOHN MATTHEWS. 1997. The role of feature geometry in the development of phonemic contrasts. In *Focus on phonological acquisition*, ed. by S. J. Hannahs and Martha Young-Scholten, 67–112. Amsterdam: John Benjamins.
- BROWN, ROGER. 1973. *A First Language: The Early Stages*. Cambridge, MA: MIT Press.
- CARAMAZZA, ALFONSO. 1988. Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Annual Review of Neuroscience* 11.395–421.

- CHAO, YUEN-REN. 1951. The Cantian idiolect: an analysis of the Chinese spoken by a twenty-eight-months-old child. *University of California Publications in Semitic Phonology* 11.27–44.
- CHARLES-LUCE, JAN, and PAUL LUCE. 1990. Some structural properties of words in young children's lexicons. *Journal of Child Language* 17.205–215.
- CHOMSKY, NOAM. 1964. Current Issues in Linguistic Theory. In *The Structure of Language: Readings in the Philosophy of Language*, ed. by Jerry Fodor and Jerrold Katz. Englewood Cliffs, NJ: Prentice Hall.
- . 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- , and MORRIS HALLE. 1968. *The Sound Pattern of English*. New York, NY: Harper and Row.
- CURRIE-HALL, KATHLEEN. 2006. Finding vowels without phonology? In *Montreal-Ottawa-Toronto Phonology Workshop*.
- DAVIS, BARBARA L, and PETER F MACNEILAGE. 1990. Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing* 33.16–27.
- DE BOER, BART, and PATRICIA K. KUHL. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online* 4.129–134.
- DILLON, BRIAN, EWAN DUNBAR, and WILLIAM IDSARDI. 2012. A single stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive Science*. In press.
- DINNSSEN, DANIEL A. 1992. Variation in Developing and Fully Developed Phonetic Inventories. In *Phonological Development: Models, Research, Implications*, ed. by Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon, 191–210. Timonium, MD: York Press.
- . 1996. Context effects in the acquisition of fricatives. In *UBC International Conference on Phonological Acquisition*, 136–148.
- , KATHLEEN M. O'CONNOR, and JUDITH A. GIERUT. 2001. The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition. *Journal of Linguistics* 37.503–525.
- DOYA, KENJI, SHIN ISHII, ALEXANDRE POUGET, and RAJESH P. N. RAO. 2007. *The Bayesian Brain: Probabilistic Approaches to Neural Coding*. Cambridge, MA: MIT Press.
- DRESHER, B. ELAN. 2009. *The Contrastive Hierarchy in Phonology*. Cambridge, UK: Cambridge Univ Pr.
- , and JONATHAN KAYE. 1990. A computational learning model for metrical phonology. *Cognition* 34.137–195.

- DUNBAR, EWAN, and WILLIAM J. IDSARDI. 2010. Review of Daniel Silverman (2006). A critical introduction to phonology: of sound, mind, and body. In *Phonology*, volume 27, 325–331.
- DYCK, CARRIE, 1995. *Constraining the phonetics-phonology interface: with exemplification from Spanish and Italian dialects*. University of Toronto dissertation.
- EDWARDS, JAN, and MARY BECKMAN. 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development* 4.122–156.
- EILERS, R E, and D. KIMBROUGH OLLER. 1976. The role of speech discrimination in developmental sound substitutions. *Journal of Child Language* 3.319–329.
- EIMAS, PETER D., EINAR R. SIQUELAND, PETER JUSCZYK, and JAMES VIGORITO. 1971. Speech perception in infants. *Science* 171.303–306.
- FELDMAN, NAOMI H, THOMAS L GRIFFITHS, and JAMES L MORGAN. 2009. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review* 116.752–782.
- FENNELL, CHRISTOPHER T, 2004. *Infant attention to phonetic detail in word forms: Knowledge and familiarity effects*. University of British Columbia dissertation.
- . 2006. Infants of 14 months use phonetic detail in novel words embedded in naming phrases. In *Proceedings of the 30th annual Boston University Conference on Language Development*. Cascadilla Press.
- , SANDRA R WAXMAN, and ADRIANA WEISLEDER. 2007. With Referential Cues, Infants Successfully Use Phonetic Detail in Word Learning. In *Proceedings of the 31st Annual Boston University Conference on Language Development*, 178–189. Cascadilla Press.
- , and JANET F WERKER. 2003. Early word learners' ability to access phonetic detail in well-known words. *Language and speech* 46.245–64.
- FENNELL, CT, and JF WERKER. 2004. Infant attention to phonetic detail: Knowledge and familiarity effects. In *Proceedings of the 28th annual Boston University conference on language development*. Cascadilla Press.
- FIKKERT, PAULA, 1994. *On the Acquisition of Prosodic Structure*. University of Leiden dissertation.
- . 2005. Getting sound structures in mind. Acquisition bridging linguistics and psychology? In *Twenty-first century psycholinguistics: Four cornerstones*, ed. by Anne Cutler, 43–56. Mahwah, NJ: Lawrence Erlbaum.
- , and CLAARTJE LEVELT. 2008. How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In *Contrast in Phonology: Theory, Perception Acquisition*, ed. by Peter Avery, B. Elan Dresher, and Keren Rice, 231–270. Berlin: Mouton de Gruyter.

- FLEMMING, EDWARD. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology*. *Phonology* 18.7–44.
- FRIEDRICH, CLAUDIA K., ADITI LAHIRI, and CARSTEN EULITZ. 2008. Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *Journal of experimental psychology: Human perception and performance* 34.1545–1559.
- GARNICA, OLGA K. 1973. The Development of Phonemic Speech Perception. In *Cognitive Development and the Acquisition of Language*, ed. by Timothy E. Moore, 215–222. New York: Academic Press.
- GIBSON, EDWARD, and KENNETH WEXLER. 1994. Triggers. *Linguistic Inquiry* 25.407–454.
- GOAD, HEATHER. 1997. Consonant Harmony in Child Language: An Optimality-Theoretic Account. In *Focus on phonological acquisition*, ed. by S. J. Hannahs and Martha Young-Scholten, 113–142. Amsterdam: John Benjamins.
- GOLDWATER, SHARON, and MARK JOHNSON. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*.
- GREENBERG, JOSEPH H. 1963. Some universals of language with special reference to the order of meaningful elements. In *Universals of Language*, ed. by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- GRIFFITHS, THOMAS L, CHARLES KEMP, and JOSHUA B TENENBAUM. 2008. Bayesian models of cognition. In *The Cambridge handbook of computational cognitive modeling*, 1–49. Cambridge, UK: Cambridge Univ Pr.
- GRUNWELL, PAMELA. 1981. The development of phonology: a descriptive profile. *First Language* 3.161–191.
- . 1982. *Clinical Phonology*. London: Croon Helm.
- HALE, MARK, and CHARLES REISS. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.
- HALLE, MORRIS, BERT VAUX, and ANDREW WOLFE. 2000. On Feature Spreading and the Representation of Place of Articulation. *Linguistic Inquiry* 31.387–444.
- HARRIS, ZELIG. 1951. *Methods in Structural Linguistics*. Chicago: Univ of Chicago Pr.
- HASTIE, TREVOR, ROB TIBSHIRANI, and JEROME FRIEDMAN. 2009. *The Elements of Statistical Learning*. New York: Springer.
- HAYES, BRUCE, and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

- INGRAM, DAVID. 1988a. Jakobson revisited: Some evidence from the acquisition of Polish. *Lingua* 75.55–82.
- . 1988b. The Acquisition Of Word-Initial [v]. *Language and Speech* 31.77–85.
- ITKONEN, TERHO. 1977. Huomioita lapsen äänteistön kehityksestä. *Virittäjä* 279–308.
- JAKOBSON, ROMAN. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksells.
- , 1968. Child Language, Aphasia, and Phonological Universals.
- , GUNNAR FANT, and MORRIS HALLE. 1952. *Preliminaries to speech analysis: The distinctive features*. Cambridge, MA: MIT Press.
- JUSCZYK, PETER. 1992. Developing Phonological Categories from the Speech Signal. In *Phonological Development: Models, Research, Implications*, ed. by Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon, 17–64. Timonium, MD: York Press.
- KAGER, RENE, JOE PATER, and WIM ZONNEVELD (eds.) 2004. *Constraints in Phonological Acquisition*. Cambridge, UK: Cambridge University Press.
- KNILL, DAVID C., and WHITMAN RICHARDS. 1996. *Perception as Bayesian Inference*. Cambridge, UK: Cambridge Univ Pr.
- KUHL, PATRICIA K., and JAMES D. MILLER. 1975. Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants. *Science* 190.69–72.
- LAHIRI, ADITI, and HENNING REETZ. 2002. Underspecified Recognition. In *Labphon 7*, ed. by Carlos Gussenhoven, Natasha Warner, and Ton Rietveld.
- LASKY, ROBERT E., ANN SYRDAL-LASKY, and ROBERT E. KLEIN. 1975. VOT Discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20.215–225.
- LEVELT, CLARA C., 1989. *An essay on child phonology*. Leiden University dissertation.
- LIEBERMAN, PHILIP. 1980. On the development of vowel production in young children. In *Child Phonology: Volume 1: Production*, ed. by Grace H. Yeni-Komshian, James F. Kavanagh, and Charles A. Ferguson, 113–142. New York: Academic Press.
- LISKER, LEIGH, and ARTHUR S. ABRAMSON. 1970. The Voicing Dimension: Some Experiments in Comparative Phonetics. In *Proceedings of Sixth International Conference of Phonetic Sciences*, 563–567.

- MACKAY, DAVID. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- MACKEN, MARLYS. 1978. Permitted Complexity in Phonological Development: One Child's Acquisition of Spanish Consonants. *Lingua* 44.219–253.
- . 1979. Developmental Reorganization of Phonology: A Hierarchy of Basic Units of Acquisition. *Lingua* 49.11–49.
- . 1980. The child's lexical representation: the 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* 16.1–17.
- , and DAVID BARTON. 1980. A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice onset time. *Journal of Child Language* 7.41–74.
- MAGRI, GIORGIO. 2012. HG has no computational advantages over OT. In *WCCFL 29: Proceedings of the 29th West Coast Conference in Formal Linguistics*, ed. by J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman, 380–388, Somerville, MA.
- MAJOR, ROY C., 1976. *Phonological Differentiation of a Bilingual Child*. Ohio State University dissertation.
- MCMURRAY, BOB, and RICHARD N ASLIN. 2005. Infants are sensitive to within-category variation in speech perception. *Cognition* 95.B15–826.
- , ———, and JOSEPH C TOSCANO. 2009. Statistical learning of phonetic categories: insights from a computational approach. *Developmental science* 12.369–78.
- MENN, LISE. 1980. Phonological Theory and Child Phonology. In *Child Phonology: Volume 1: Production*, ed. by Grace H. Yeni-Komshian, James F. Kavanah, and Charles A. Ferguson, 23–42. New York: Academic Press.
- . 1992. Building Our Own Models: Developmental Phonology Comes of Age. In *Phonological Development: Models, Research, Implications*, ed. by Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon, 3–15. Timonium, MD: York Press.
- NÄÄTÄNEN, R, A.W.K. GAILLARD, and S. MÄÄNTYSALO. 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42.313—329.
- NIYOGI, P, and R C BERWICK. 1996. A language learning model for finite parameter spaces. *Cognition* 61.161–93.
- OLMSTED, DAVID L. 1971. *Out of the Mouth of Babes: Earliest Stages in Language Learning*. The Hague: Mouton De Gruyter.
- OTOMO, K, and C STOEL-GAMMON. 1992. The acquisition of unrounded vowels in English. *Journal of speech and hearing research* 35.604–16.

- PALLIER, CHRISTOPHE, NURIA SEBASTIÁN-GALLÉS, and ANGELS COLOMÉ. 1999. Phonological representations and repetition priming. In *Eurospeech '99*, 1907–1910.
- PATER, JOE. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39.334–345.
- , CHRISTINE STAGER, and JANET WERKER. 2004. The Perceptual Acquisition of Phonological Contrasts. *Language* 80.384–402.
- , CHRISTINE STAGER, and JANET F. WERKER. 1998. Additive effects of phonetic distinctions in word learning. In *Proceedings: 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America.*, 2049–2050.
- PAVLOVIĆ, MILIVOJ. 1920. *Le langage enfantin: Acquisition du Serbe et du Français par un enfant serbe*. Paris: E. Champion.
- PEPERKAMP, SHARON, ROZENN LE CALVEZ, JEAN-PIERRE NADAL, and EM-MANUEL DUPOUX. 2006. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101.B31–841.
- PIERREHUMBERT, JANET B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46.115–154.
- PRATHER, ELIZABETH M., DONA LEE HEDRICK, and CAROLYN A. KERN. 1975. Articulation Development in Children Aged Two to Four Years. *Journal of Speech and Hearing Disorders* 40.170–191.
- PRINCE, ALAN, and PAUL SMOLENSKY. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- PYE, CLIFTON, DAVID INGRAM, and HELEN LIST. 1987. A comparison of initial consonant acquisition in English and Quiche. *Children's Language* 6.175–190.
- RICE, KEREN, and PETER AVERY. 1995. Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In *Phonological acquisition and phonological theory*, ed. by John Archibald, 23–42. Hillsdale, NJ: Lawrence Erlbaum.
- SCHARINGER, MATHIAS, PHILIP J MONAHAN, and WILLIAM J IDSARDI. 2012. Asymmetries in the Processing of Vowel Height. *Journal of Speech, Language and Hearing Research* 55.903–919.
- SCOBIE, JAMES M., FIONA GIBBON, WILLIAM J. HARDCASTLE, and PAUL FLETCHER. 2000. Covert contrast as a stage in the acquisition of phonetics and phonology. In *Papers in laboratory phonology V: Language acquisition and the lexicon*, ed. by Michael B. Broe and Janet B. Pierrehumbert, 194–207. Cambridge, UK: Cambridge Univ Pr.
- SHVACHKIN, N. 1948. The Development of Phonemic Speech Perception in Early Childhood. In *Studies of Child Language Development*, ed. by Charles A. Ferguson and Dan I. Slobin, 91–127. New York: Holt, Rinehart and Winston.

- SIQUELAND, EINAR R., and CLEMENT A. DELUCIA. 1969. Visual Reinforcement of Nonnutritive Sucking in Human Infants. *Science* 165.1144–1146.
- SLEDD, JAMES H. 1966. Breaking, Umlaut, and the Southern Drawl. *Language* 42.18–41.
- SMITH, NEIL. 1973. *The Acquisition of Phonology*. Cambridge, UK: Cambridge Univ Pr.
- . 2010. *Acquiring Phonology: A Cross-Generational Case Study*. Cambridge, UK: Cambridge Univ Pr.
- STAGER, C L, and J F WERKER. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388.381–2.
- STAMPE, DAVID. 1969. The Acquisition of Phonetic Representations. In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, 443–454.
- STRAIGHT, H. STEPHEN. 1980. Auditory versus Articulatory Phonological Processes and Their Development in Children. In *Child Phonology: Volume 1: Production*, ed. by Grace H. Yeni-Komshian, James F. Kavanagh, and Charles A. Ferguson, 43–67. New York: Academic Press.
- STREETER, LYNN A. 1976. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259.39–41.
- SWINGLEY, DANIEL, and RICHARD N. ASLIN. 2000. Spoken word recognition and lexical representation in very young children. *Cognition* 76.147–166.
- , and ———. 2002. Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychological Science* 13.480–484.
- TENENBAUM, JOSHUA, 1999. *A Bayesian Framework for Concept Learning*. Cambridge, MA: MIT dissertation.
- TESAR, BRUCE, and PAUL SMOLENSKY. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29.229–268.
- THIESSEN, ERIK D. 2007. The effect of distributional information on children’s use of phonemic contrasts. *Journal of Memory and Language* 56.16–34.
- VALLABHA, GAUTAM K, JAMES L MCCLELLAND, FERRAN PONS, JANET F WERKER, and SHIGEAKI AMANO. 2007. Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America* 104.13273–8.
- VANDERWEIDE, TERESA. 2006. Cues, Opacity and the puzzle-puddle-pickle problem. In *Proceedings of the 2006 Canadian Linguistics Association Annual Conference*, 115–130. Citeseer.

- VIHMAN, MARILYN MAY, and WILLIAM CROFT. 2007. Phonological development: Toward a 'radical' templatic phonology. *Linguistics* 45.683–725.
- , CHARLES A. FERGUSON, and MARY ELBERT. 1986. Phonological Development From Babbling to Speech: Common Tendencies and Individual Differences. *Applied Psycholinguistics* 7.3–40.
- WALES, JULIA, and GEORGE J. HOLLICH. 2004. Early word learning: How infants learn words that sound similar. In *Cognitive Science Society*.
- WERKER, JANET F. 1989. Becoming a native listener. *American Scientist* 77.54–59.
- , LESLIE B. COHEN, VALERIE L. LLOYD, MARIANELLA CASASOLA, and CHRISTINE L. STAGER. 1998. Acquisition of Word-Object Associations by 14-Month-Old Infants. *Developmental Psychology* 34.1289–1309.
- WERKER, JANET F, and CHRISTOPHER E FENNELL. 2004. From listening to sounds to listening to words: Early steps in word learning. In *Weaving a Lexicon*, ed. by D. Geoffrey Hall and Sandra R. Waxman, 79–110. Cambridge, MA: MIT Press.
- WERKER, JANET F., CHRISTOPHER T. FENNELL, KATHLEEN M. CORCORAN, and CHRISTINE L. STAGER. 2002. Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy* 3.1–30.
- , and RICHARD C. TEES. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7.49–63.
- YANG, CHARLES. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.
- YOSHIDA, KATHERINE A., CHRISTOPHER T. FENNELL, DANIEL SWINGLEY, and JANET F. WERKER. 2009. Fourteen-month-old infants learn similar-sounding words. *Developmental science* 12.412–418.