Review TRENDS in Cognitive Sciences Vol.9 No.11 November 2005

Infant speech perception bootstraps word learning

Janet F. Werker and H. Henny Yeung

Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver BC, V6T 1Z4, Canada

By their first birthday, infants can understand many spoken words. Research in cognitive development has long focused on the conceptual changes that accompany word learning, but learning new words also entails perceptual sophistication. Several developmental steps are required as infants learn to segment, identify and represent the phonetic forms of spoken words, and map those word forms to different concepts. We review recent research on how infants' perceptual systems unfold in the service of word learning, from initial sensitivity for speech to the learning of language-specific sound patterns. Building on a recent theoretical framework and emerging new methodologies, we show how speech perception is crucial for word learning, and suggest that it bootstraps the development of a separate but parallel phonological system that links sound to meaning.

Introduction

Only humans acquire language. Perhaps this is why the first words learned by infants seem so special – word learning is a milestone on the path towards developing a uniquely human ability. But the task of word learning, beginning with recognizing spoken words, is not trivial. It requires a complex mapping among a concept, a word, and the word's corresponding acoustic signal across different speakers and phonetic contexts. Although a long tradition in infancy research investigates how conceptual systems develop and then change as words are learned, it is only recently that researchers have begun to understand the vital role infants' developing *perceptual* systems play in word learning.

Previously, it had been assumed that the perceptual units required for lexical acquisition were available as representations that could simply be mapped onto their corresponding concepts. It is now known that the emergence of perceptual units for lexical acquisition has a developmental history, and that the same processes that shape these units simultaneously enable the acquisition of other, grammatical properties of the language. Moreover, very recent work suggests that these perceptual units continue to change throughout the early stages of word learning. This review highlights these recent empirical findings, many using emerging technologies, which reveal how perceptual systems for speech are (i) initially structured, (ii) change with language-specific exposure, and then (iii) contribute to, and are changed further by, the process of word learning.

A theoretical launching point for this review is the notion of 'bootstrapping' – using existing knowledge to facilitate acquisition of novel abilities. We begin with the idea that the perceptual biases infants have at birth serve as the 'primitives' from which word forms are constructed. General perceptual learning enables infants to extract these early word forms, using increasingly precise knowledge of the acoustic and phonetic properties of the native language. These word forms are fragile in early development, but once they are linked with concepts, a stable phonological representation of word forms is *bootstrapped* from the existing perceptual system. This phonological system is what enables efficient extraction, maintenance and linkage of word representations to concepts by 18–20 months.

Initial perceptual biases

Neonates show several perceptual biases, some of which vary as a function of prenatal exposure. They prefer their mother's voice, stories and songs heard prenatally, and their native language [1]. Even fetuses appear to show preference, as measured by heart-rate, for their mother's voice [2]. These reports confirm that prenatal auditory experience tunes neonatal perception.

Other early-emerging perceptual sensitivities are more difficult to explain through prenatal learning and probably reflect either general properties of animal auditory systems or epigenetically determined, uniquely human biases. Sensitivities shared with non-human animals include neonates' ability to discriminate languages with different rhythmical properties only when speech is played forwards, not backwards [3], and early-appearing categorical-like discrimination of phonetic contrasts ([4], but see [5]). Further research is needed to determine if all initial biases in infants are shared with other animals, including the preference for speech over acoustically matched nonspeech [6], discrimination of lexical *versus* grammatical words [7], and sensitivity to phonetic cues that indicate word boundaries [8]. These initial biases and capabilities, irrespective of their origins, prepare the perceptual system for later speech input from the environment.

Cognitive neuroscience complements behavioral work, showing that initial neural organization in neonates has some specificity for speech signals, but requires further

Corresponding author: Werker, J.F. (jwerker@psych.ubc.ca).

Available online 3 October 2005

www.sciencedirect.com 1364-6613/$ - see front matter doi:10.1016/j.tics.2005.09.003

experience to establish adult-like organization. Imaging studies reveal unique cortical activation to speech in comparison with equally complex backward speech (e.g. [9,10]) and electrophysiological studies report unique neural activity to changes in phonetic *versus* non-phonetic attributes [11]. Neural organization is further refined, with adult-like left hemisphere dominance for speech, by 10–12 months [11]. However, limited plasticity is also seen. Damage to the left hemisphere in infancy can lead to reversal of dominance, with similar areas in the right hemisphere taking over the phonetic tasks [12].

Neural and perceptual systems are initially organized to treat speech sounds differently from non-speech, but are also dependent on experience (see Box 1 for further discussion about lasting impacts of early experience). The next section describes how native language input further refines perceptual sensitivities for speech, eventually helping infants attend to, segment and remember words.

Language-specific reorganization

Native language input acts to *reorganize* perceptual sensitivity, selectively maximizing attention to phonetic features that distinguish native language categories. For example, infants begin life discriminating both native and non-native phonetic contrasts, but by 6–12 months of age show a decline in discrimination of many non-native distinctions and an enhancement of sensitivity to native ones [13,14]. The timing and extent of this reorganization is influenced by several factors, including the acoustic/articulatory characteristics of the phonetic contrast [15] and the similarity of the contrast to those used in the native language [16]. Within the same time period, infants learn many other phonological properties of the native language; as reviewed by Jusczyk [17], infants by 9–10 months prefer well-formed words that correspond to frequent patterns in the input.

Changes in perceptual sensitivity that occur in the first year of life have been referred to as a ‘functional reorganization’, a term that describes developing patterns of discrimination in accordance with functional categories in the native language, but does not imply loss of perceptual ability [14,18]. With sensitive measures, for example, both adults [19] and infants [20] respond to phonetic differences that are not contrastive in the native language.

Box 1. Assessing the impact of early exposure

Both longitudinal studies and studies with bilingual and special populations provide evidence linking early perceptual sensitivity to language proficiency. Individual performance in vowel discrimination tasks at 6 months predicts vocabulary size, as well as scores on other language measures from 13–24 months of age [62]. Reading proficiency in children 3 to 8 years of age is also correlated with electrophysiological measures of phonetic discrimination recorded when these individuals were neonates [63]. Although these studies do not assess language-specific perceptual learning *per se*, they provide strong evidence that early general perceptual sensitivities are correlated with proficiency in later language acquisition.

Evidence for the importance of early perceptual experience also comes from studies involving infants who experience partial or total hearing loss. Infants with a history of middle ear infections are more likely to have later language delay [64]. Moreover, infants born deaf and then fitted with cochlear implants at 17–24 months (and tested 2–18 months later) recover the ability to discriminate phonetic distinctions, but remain compromised in their ability to make word–object associations [65] in comparison with infants who had implants from 7–15 months. These studies provide intriguing, but preliminary evidence that there is a sensitive period in which perceptual exposure is necessary for subsequent facile word learning.

Studies of adult second-language learners suggest that a lack of not just auditory exposure in general, but also early language-specific exposure has long-term consequences for both the lexical use of phonetic contrasts and for phonetic perception. For example, highly proficient Spanish–Catalan bilinguals, who first learned Catalan at 3–4 years of age are less able to use Catalan-specific distinctions in lexical decision tasks than native Catalan bilinguals [66]. Moreover, they are not as proficient at discriminating Catalan-specific vowel distinctions [67].

Early exposure might lead to lasting effects only if there is at least some continuing exposure. Adults learning Korean up to 3–8 years of age, and then adopted into French homes without any Korean exposure were no better able to discriminate Korean-specific phonetic distinctions than French adults [68]. However, second-language learners of either Korean or Spanish who overheard either language before age 5, and then were exposed for just a few hours a week throughout childhood, were able to maintain native-like discrimination for Korean phonetic contrasts [69] and production for Spanish contrasts [70], whereas learners without this early and continued exposure performed significantly worse.

Statistical learning: a mechanism for reorganization?

One mechanism that underlies functional reorganization might be statistical – as infants are exposed to language-specific input, emergent properties of the input may shape the perceptual system. By at least 9 months, infants are sensitive to the frequency, distribution, and other statistical properties of perceptual input in speech [13]. Highly frequent phonetic contrasts and phonotactic patterns (i.e. permissible combinations of sounds) are categorized in a language-specific manner at younger ages than less frequent ones [13,21]. After repeated exposure to lists of nonsense words with recurring sound patterns, infants can make generalizations about syllable structure [22], stress [23] and phonotactic patterns [24].

In phonetic perception, frequent exemplars define the centers of perceptual categories. As illustrated in Figure 1, simply changing the frequency distribution of the input can lead to a modification in phonetic categories in infants 6–8 months of age [25]. Frequency detection also underlies the perceptual magnet effect, where central exemplars serve to attract other members of the category, thus diminishing discrimination within a category [26]. Indeed distributional input might drive functional reorganization by shrinking and expanding the perceptual distances within and between categories [27].

Statistical learning requires only attention and exposure to input, but as infants mature, they have access to more sophisticated cognitive abilities. One recent study suggests that contingent social interaction but not simple exposure changes phonetic discrimination after 9–10 months [28]. Future research must investigate whether the mechanisms for functional reorganization change across development.

Perceptual basis of word learning

In addition to the conceptual barriers that infants overcome before they learn to use words appropriately, infants

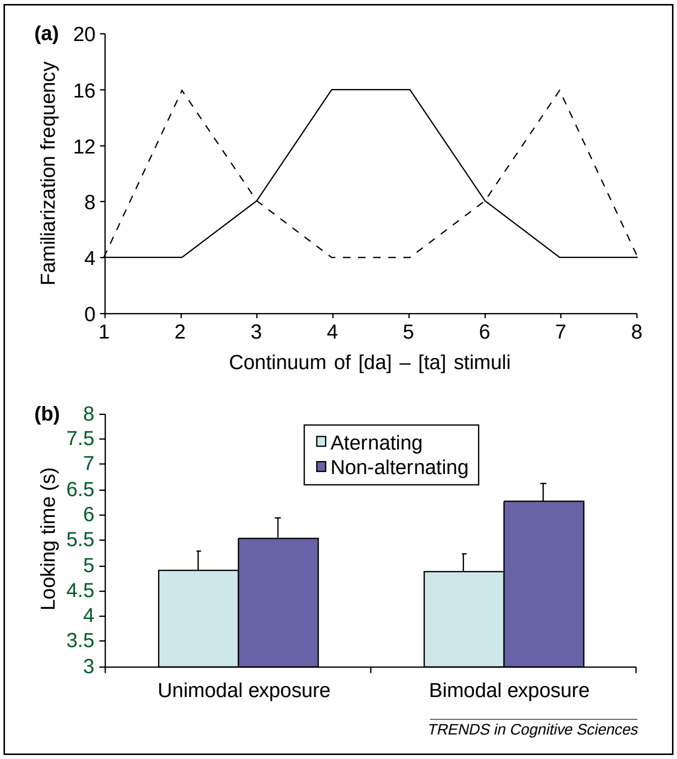

Figure 1. (a) English-learning infants were exposed to an 8-step continuum of stimuli modeled on a phonetic difference that is not used contrastively in English (voiced and unaspirated voiceless alveolar obstruents [da] and [ta]). For 2.3 min, these stimuli were presented in either a Bimodal (dotted line) or a Unimodal (solid line) frequency distribution [25]. **(b)** Following exposure, infants were presented with *alternating* (both stimuli 1 and 8) and *non-alternating* (either stimulus 3 or 6) test trials. Infants in the Bimodal condition discriminated between non-alternating and alternating trials, whereas infants in the Unimodal condition listened equally to both. Data from [25].

face perceptual challenges: they must also learn to recognize and represent contrasting acoustic forms of words. Here our discussion is restricted to how infants (i) segment words from continuous speech, (ii) remember words as distinct from one another, and (iii) begin to map those word forms onto referents.

Word segmentation

Although standard theories of language acquisition once began with the assumption that words are perceptually available from the beginning, word boundaries are not acoustically demarcated in continuous speech. More than 10 years of research has looked at how infants segment words from the speech stream without *a priori* knowledge of word forms. Jusczyk and Aslin first demonstrated that infants begin to segment words by 7–8 months of age; after being familiarized with words such as '*cup*' and '*dog*', infants listen longer to passages containing those words over passages containing other equally common words [29]. At 7 months, English-learning infants pull out words that conform to the common English strong-weak stress pattern, like '*DOCtor*', but do not segment weak-strong words like '*guiTAR*'. By 10 months, English-learning infants can also segment weak-strong words, perhaps because they can also use language-specific phonetic and phonotactic cues to word boundaries [17,30]. All of these cues improve performance in computational models of word segmentation [31]. Once learned, frequent word forms, like the infant's name, facilitate segmentation of new words [32].

Another statistical regularity that infants are sensitive to is 'transitional probability', learning that syllables from within one word are more likely co-occur than syllables from separate words [13]. An emerging debate is whether infants first begin to segment words by using transitional probabilities [33], or by using word-level, native-language phonological properties, such as strong–weak stress patterns for English words, learned initially from words presented in isolation [34,35].

Word forms

By 9–10 months of age infants show an increasing preference for word forms that conform to the phonological characteristics of the native language. Infants of 9–10 months prefer to listen to words obeying native language phonotactics and to words with native language stress patterns (see [17,30] for reviews).

In addition to language-specific constraints on word forms, infants also encode phonetic detail (e.g. '*tup*' is not confused with '*cup*' [29]) and indexical detail (such as speaker identity [36] and emotional affect [37]). Infants fail to recognize repetitions of a word, particularly after some delay, if the indexical properties are changed. By 10–11 months, infants *are* able to recognize the word form across these indexical changes, as well as when syllabic stress changes [38], but recognition of these word forms is still faster when indexical information remains constant [36].

Although these results suggest that infants learn to give more weight to phonetic detail in word forms by 10–11 months, access is still fragile. Infants of this age treat mispronounced words like real words, although only when these mispronunciations are perceptually confusable (Table 1). Infants listening to pseudowords like '*didder*' treat them as real words like '*dinner*' because they differ on unstressed syllables [38,39]. However, infants show inconsistent treatment of pseudowords that differ on stressed syllables that are not in word-initial position [39] and that differ in syllable-final positions [40]. Pseudowords like '*ninner*' (similar to '*dinner*') and '*pog*' (similar to '*dog*') are treated like unknown words, perhaps because these words differ in syllable-initial position [40] on perceptually prominent stressed syllables [38].

Pairing words and objects

Infants begin with simple associations between words and objects. By 6 months of age, for example, infants associate highly frequent words, such as '*Mommy*', and their referents [41]. Over the next 8 months, infants develop cognitive and perceptual abilities that allow learning of new associations more quickly, and in increasingly unconstrained situations. Infants as young as 8 months are able to link novel words to novel objects after only a few repetitions of the pairing, but require cross-modal synchrony between the presentation of the word and movement of the object [42]. Learning associative links at 12 months still relies heavily on perceptual and social cues like visual salience and eye-gaze; for example, infants think an attractive object is being labeled even when

Table 1. Developing access to phonetic detail as infants progress from word form processing to word learning

Task	Conclusions	Stimuli examples	Phonetic transcription[a]	Age (months) 7–9	10–12	13–15	17–24
Word forms							
Discrimination of novel word forms	Phonetic sensitivity at 8 months.	bih / dih [47]	[bɪ] / [dɪ]	✓			
Recognizing familiar word forms	At 9-months, indexical/ stress information is weighed as heavily as phonetic information, but by 10–12 months, phonetic information is weighed more heavily	cup / tup [29]	[kʰʌp] / [tʰʌp]	✓			
		cup / *cup* [b] [37]	[kʰʌp] / [*kʰʌp*]	✓	×		
		cup$_1$ / cup$_2$ [c] [36]	[kʰʌp]$_1$ / [kʰʌp]$_2$	✓	×		
		diNNER / DInner [d] [38]	[dɪ.ˈnɚ] / [ˈdɪ.nɚ]		×		
	At 10–12 months, use of phonetic information still depends on perceptual salience. Task is easier when critical phonetic detail is in stressed, word-initial, and/or syllable-initial positions	DIdder / DInner [d] [38]	[ˈdɪ.dɚ] / [ˈdɪ.nɚ]		×		
		bonJOUR / ponJOUR [e] [39]	[bɔ̃.ˈʒuʀ] / [pɔ̃.ˈʒuʀ]		×		
		bonJOUR / bonGOUR [d,e] [39]	[bɔ̃.ˈʒuʀ] / [bɔ̃.ˈguʀ]		? [f]		
		paart / paarp [g] [40]	[pʰaːrt] / [pʰaːrp]		? [f]		
		paart / daart [g] [40]	[pʰaːrt] / [daːrt]		✓		
		DInner / NInner [d] [38]	[ˈdɪ.nɚ] / [ˈnɪ.nɚ]		✓		
Word-object pairings							
Learning novel word–object pairings[h]	Learning of novel word–object pairings by 14 months, but phonetic detail not accessed until 17 months	lif / neem [44,45]	[lif] / [nim]	×	×	✓	
		pin / din [46]	[pʰɪn] / [dɪn]			×	
		bih / dih [47]	[bɪ] / [dɪ]			×	✓
	Phonetic detail accessed in easier tasks — Audiovisual synchrony	tah / gah [42]	[tʰa] / [ga]	✓			
	Word forms familiar	ball / doll [50]	[bɑl] / [dɑl]			✓	
	Preferential Looking Task	tuk / duk [49]	[tʰʌk] / [dʌk]			✓	
Recognizing familiar word–object pairings[i]	From 14–24 months, phonetic detail accessed in word recognition	ball / doll [49,72]	[bɑl] / [dɑl]			✓	✓
		baby / vaby [51,71]	[ˈbeɪ.bi] / [ˈveɪ.bi]			✓	✓

✓= stimuli pairs treated as distinct word forms. ×= not treated as distinct word forms in task.

[a]Phonetically transcribed using International Phonetic Alphabet (IPA) characters. [b]Emotional affect was assessed (e.g. 'cup' in happy affect *versus* in normal affect). [c]Talker-change was assessed (e.g. 'cup' said by one talker *versus* by another talker). [d]These pairs of stimuli were not directly compared. However, the design allowed interpretation as if there had been a direct comparison. See [38–40] for details. [e]These experiments were carried out in France, and so 'bonjour' (hello) was assumed to be a familiar word. [f]Interpretation of results depend on task and analysis. See [38–40] for details. [g]These experiments were carried out in the Netherlands, and so 'paart' (horse) was assumed to be a familiar word. [h]Assessed primarily by using the Switch Task for word learning (see Box 2). [i]Assessed by using the Preferential Looking Task (see Box 2).

experimenter eye-gaze is directed at another object [43]. The ability to form word–object links on the basis of co-occurrence alone, without facilitating social or temporal cues is evident by 13–15 months in laboratory tasks [44,45].

At 14 months, infants' ability to associate novel words with novel objects is still dependent on the contrastive saliency of the words themselves. For example, in the 'Switch' procedure (Box 2), infants this age reliably learn to associate words such as '*lif*' and '*neem*' with two different objects. Yet, when tested with minimally contrastive words such as '*bin*' and '*pin*', 14-month-old infants fail [46–48]. This failure seems paradoxical, because 8 to 14-month-old infants can still discriminate the same two word forms when no referent is attached [47].

Why do infants at 14 months confuse phonetically similar words when they are linked to objects, yet discriminate those same words when not paired with objects? The perceptual sensitivity needed to make fine distinctions exists, but access may be inhibited by the computational demands of having to link word forms and objects [48]. Both changes to the testing conditions [49] and the use of familiar words [50] ease the processing load, enabling access to phonetic detail at 14 months. Moreover, in word recognition tasks (Box 2), where new associations do not need to be formed, infants of 14 months look longer to a target picture, such as a baby, when the target word is pronounced correctly ('*baby*') than when it is mispronounced (e.g. '*vaby*') [51].

Box 2. Assessing infants' knowledge and learning of words

The 'Switch' task, as shown in Figure Ia, assesses the ability to make novel word–object associations. Infants are habituated to two word–object combinations, where each object is presented visually while a word is played from an audio speaker. During each trial, one of the two words is presented 7–10 times while the object moves back and forth on a screen. Infants are presented with both word–object pairings until they habituate to the stimuli – that is, until their looking times to the objects decrease by a preset criterion. Following habituation, infants are tested on two types of trials. A Same trial involves a familiar word and familiar object in a familiar pairing. A Switch trial involves a familiar word and familiar object, but with the familiar pairing violated. If the infant has learned the word, the object, and their link, she should be surprised and look longer at the Switch than at the Same trial [45].

In the Preferential Looking Word Recognition procedure, recognition of known words is assessed. As shown in Figure Ib, infants are shown two pictures side-by-side, only one of which corresponds to a word embedded in a carrier phrase (e.g. 'Where's the ——?') presented over a loudspeaker. The infant's overall looking time to the match and to the mismatch can be measured, but more often looking time is only recorded during a time window of around 350–2000 ms after the onset of the spoken word. This task also allows for a measure of on-line processing by monitoring eye movements during the trials. The latency for looking away from the mismatch compared with the match can provide a sensitive index of just when it is that infants have finished processing the word and are able to detect the match [71,72].

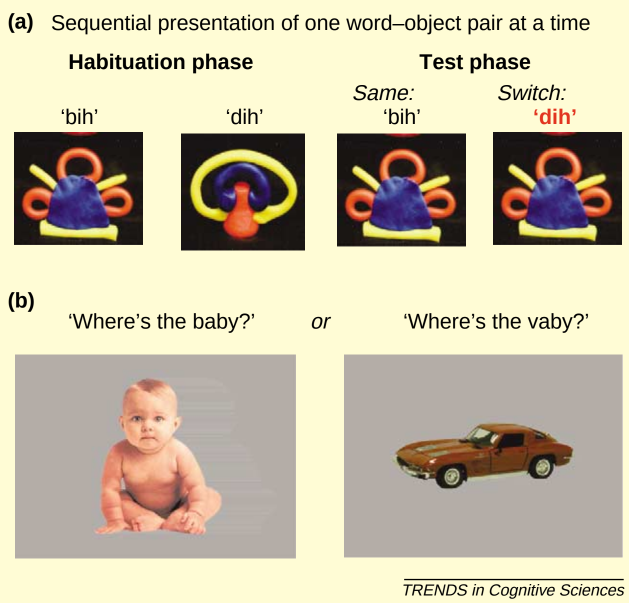

Figure I. (a) The Switch task used to measure infants' learning of novel word–object associative pairings. **(b)** The Preferential Looking Word Recognition procedure for measuring infants' word recognition abilities.

By 17 months, infants regain access to phonetic sensitivity when learning novel pairings, mapping forms like '*bih*' and '*dih*' onto two different objects (see [48]).

Neural representations of word forms

Recent methodological advancements have shown how neural representations of word forms can be studied in infancy. In 14-month-old infants, paradigms measuring event-related potentials (ERPs) from auditory words that match simultaneously presented visual pictures diverge from those of mismatched words 400–800 ms after word-onset [52], as can be seen in Figure 2a. The polarity, latency and distribution of this component suggest that it is a precursor to the well-studied N400 component in adults elicited when semantic incongruities are presented.

In paradigms where 11 to 20-month-old infants are presented with only auditory stimuli [53], differences in evoked brain potentials between known and unknown words are observed as early as 200–400 ms after word-onset, as can be seen in Figure 2b. Notably, the distribution of this component seems to change from a bilaterally distributed one at 13 months, to a left-hemisphere dominant one at 20 months [54].

As shown in Figure 3, this N200–400 component reveals further development change in word representations from 14 to 20 months [53]. Just as correctly pronounced words, like '*cup*' elicit this recognition component, mispronounced words like '*tup*' *also* elicit this component at 14 months, but not at 20 months of age. These data complement behavioral evidence that there is a change between 14 and 20 months in infants' ability to access phonetic detail in lexical tasks.

Implications for development

Linguistic bases of perceptual learning

Ultimately it is speech, and not other sounds, that is used as the medium for spoken language. Early-appearing

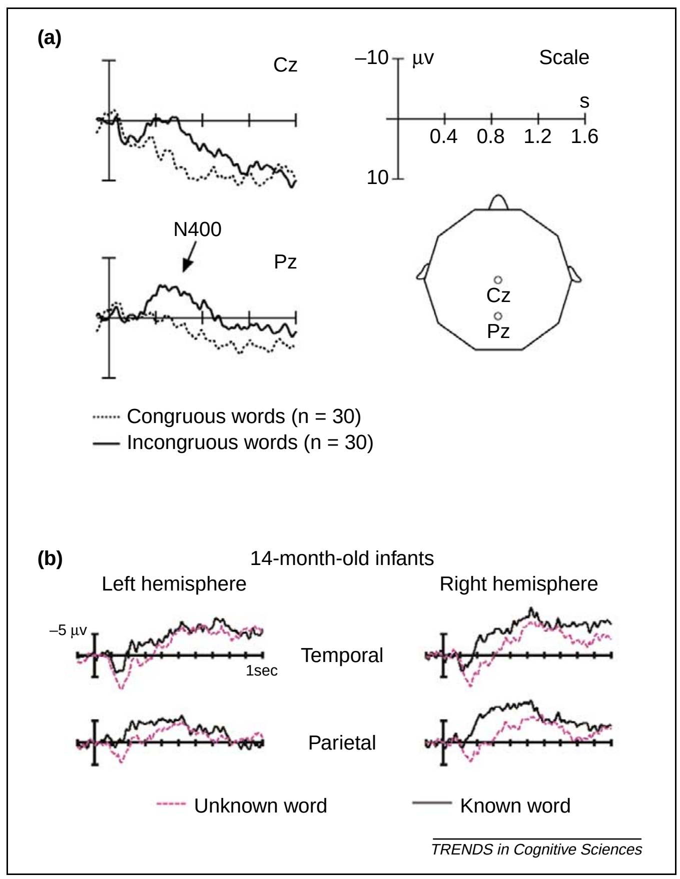

Figure 2. (a) Electrical brain activity elicited from infants aged 14 months reveals a higher negative deflection at around 400ms to known words that are congruent *versus* divergent with a simultaneously presented visual display. Figures are modified from [52], showing only central (Cz) and posterior (Pz) electrode sites. **(b)** Even when infants are just listening to words, evoked potentials from known *versus* unknown words diverge around 200–400 ms after word-onset. This ERP component indexes word recognition. Figures are modified from [53].

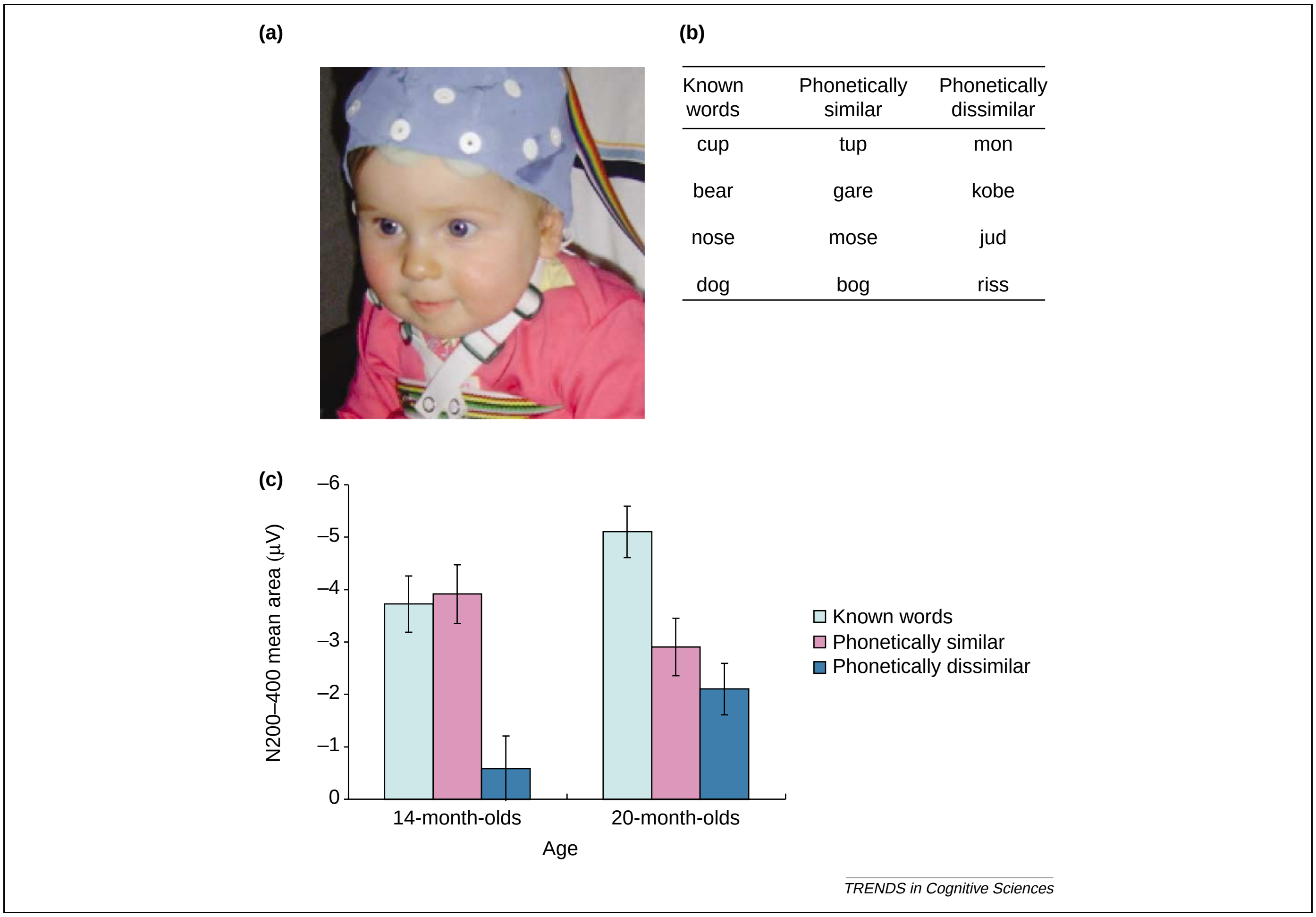

Figure 3. Infants in an ERP paradigm **(a)** were presented aurally with a list of known words and nonsense words that were either phonetically similar or dissimilar to those known words. **(b)** shows an abbreviated list. **(c)** The mean amplitude of the N200–400 word recognition component is shown in response to known words, phonetically dissimilar nonsense words, and phonetically similar nonsense words. At 20, but not 14 months, the neural representations accessed for known words are phonetically detailed enough to distinguish similar-sounding foils (c). Data are from [53].

behavioral preference [6] and unique cortical activation [9,10] for speech might give words a privileged status over non-linguistic sounds for linguistic function. At the onset of word-learning, word forms, but not tones, act as cues for 9 to12-month-old infants to individuate [55] and categorize objects [56].

Whereas conceptual systems are engaged by speech sounds early on, the meaningful application of infants' perceptual sensitivities to native language follows a somewhat different course of development. These sensitivities are not harnessed in word recognition until 14 months and in word learning tasks until 17 months (see Table 1). What developmental changes allow 17-month-old infants to learn mappings for minimally contrastive words, a skill that coincides with the beginning of the vocabulary spurt around 18 months [57]? One developmental change is learning which aspects of the acoustic signal are *functionally*, not just perceptually, distinctive. A recently advanced framework for linking speech perception to word learning, PRIMIR (Processing Rich Information from Multidimensional Interactive Representations), helps account for these data. PRIMIR suggests that although infants continue to perceive phonetic and indexical variation in word forms, by 17 months they have learned a critical number of word–object mappings. The word–object pairings highlight phonetic differences used to distinguish meaning, and allow emergence of functionally contrastive phonological categories [14]. Analogous to evidence showing that increased perceptual salience of objects makes it easier for those objects to be mapped in word-learning tasks [43], phonological categories increase perceptual salience of certain phonetic contrasts, reducing processing load and enabling efficient formation of new word–object links.

Figure 4 illustrates infants' use of phonetic detail as they process words in real-time. Speed of recognition of already-known words increases between 15 and 24 months [58], and by 18 months of age, infants recognize known words such as '*baby*' on the basis of only the initial consonant and vowel [59]. An important area for future research is to investigate whether these improvements in on-line word recognition are made possible by the emergence of functional categories.

Bootstrapping revisited

For over 30 years, researchers have asked what developments in early infancy allow word learning to proceed so rapidly in the few months before 2 years of age. Infants begin life with perceptual biases that facilitate attention to speech and the encoding of its properties. Over the first

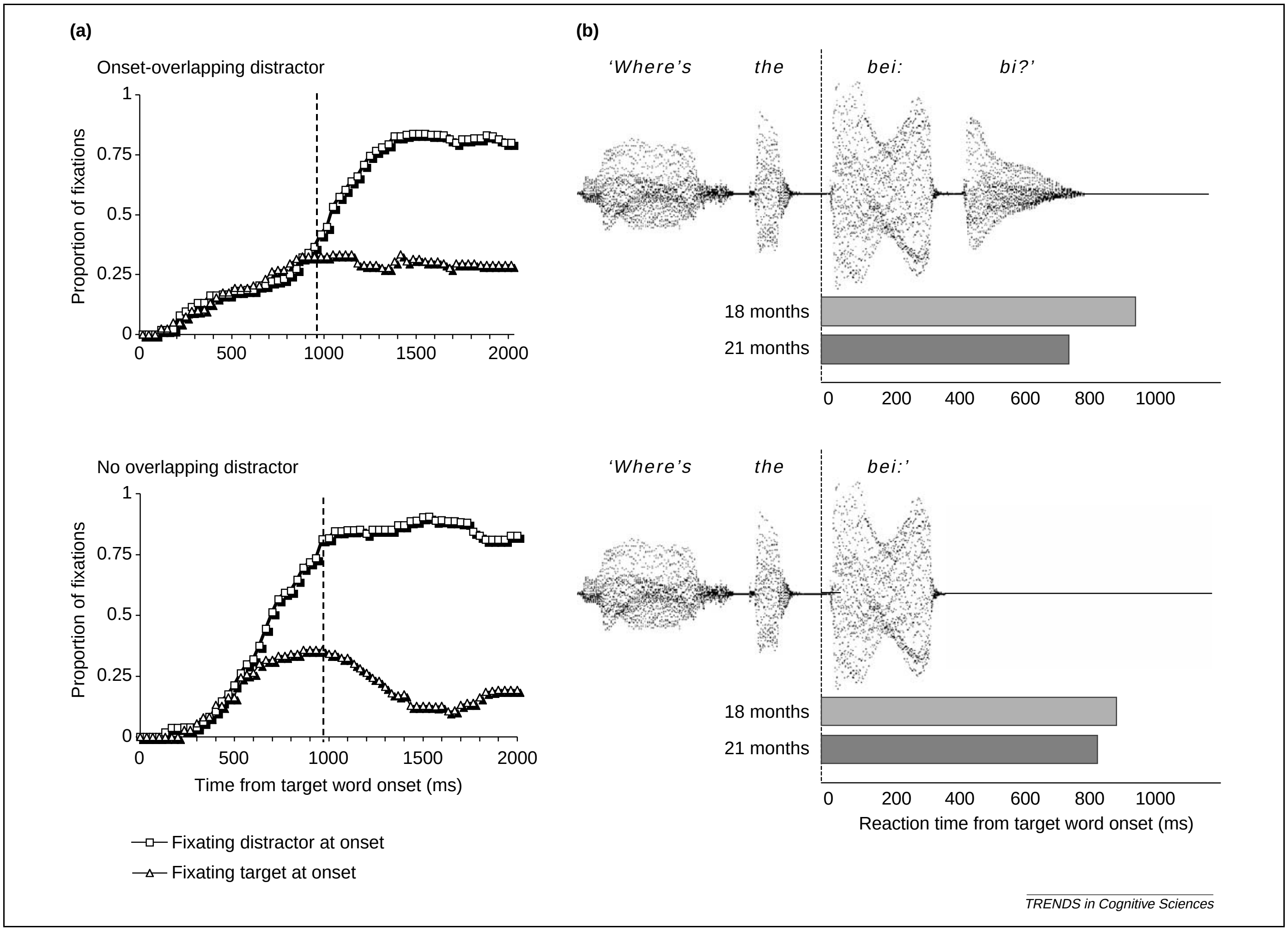

Figure 4. (a) Infants aged 24 months were shown objects corresponding to two known words that either overlapped in the initial consonant and vowel (e.g. 'dog' and 'doll') or had no overlap (e.g. 'dog' and 'tree'). Infants who were initially looking at the distractor shift to the target almost 300ms faster in the no-overlap than in the overlap trials, indicating incremental word processing. Figure adapted from [73]. **(b)** Infants aged 18 and 21 months of age were presented with either a full word (e.g. 'baby') or a partial word (e.g. 'bei'). Latency to shift to the target is evident immediately after the end of the full word, and equally rapidly when only the partial word is presented, revealing that infants initiated eye movements before the end of the word. Figure adapted from [59].

several months of life, infants' perceptual biases increasingly conform to native language patterns. Detection of statistical regularities in the input is one mechanism by which these sensitivities change, but further research (Box 3) is needed to determine if learning mechanisms change across development. Emerging native language perceptual sensitivity aids segmentation and memory of word forms, yet remains difficult to access in the initial stages of associative word learning, resulting in a U-shaped performance through early development. With the mapping of word forms onto concepts, phonological categories are bootstrapped, yielding a substantive change in the efficiency of word learning.

In summary, we suggest that word learning is another 'bootstrapping' phenomenon in developmental research. We do not suggest that word learning can be reduced to perceptual and statistical learning. Instead, we argue that perceptual learning provides a foundation upon which

Box 3. Questions for future research

- Infants are able to track the distribution of a single phonetic feature when this feature is controlled in the artificial language studies, but in natural language, the distribution of any particular feature covaries with the distribution of several other features. Can infants track the distribution of multiple cues?
- How robust is statistical learning throughout development? Do communicative intent, semantic knowledge, and/or linguistic rules, once in place, replace statistical learning as the primary engines not only of language acquisition, but also of language-specific perceptual change?
- Speech perception in infancy has focused on stressed, syllable-initial positions. Is the development of phonological knowledge in less-salient positions (like syllable-final and unstressed positions) parallel? How would this affect learning of morphological affixes, like plural '-*s*' in English?
- What are the neural mechanisms underlying changes in speech perception, detection of language-specific word forms, and the mapping of word forms to objects? Do any of these processes and their resulting representations involve specialized brain systems or do they rely on domain-general networks?
- Will an understanding of how speech perception bootstraps language acquisition address difficult theoretical issues in bilingual acquisition, such as when and if the bilingual child has one or two language acquisition systems [74]?
- Can the bootstrapping approach direct research on the early identification of language delay?

abstract linguistic units can be built. Just as phonological patterns act as cues to morphological and syntactic structure [60], and just as naïve concepts allow infants to learn more complex ones [61], perceptual learning allows segmentation and representation of word forms that, once mapped to concepts, bootstrap the process of word learning and lead to a qualitative improvement in its efficiency.

Acknowledgements

We thank Laurel Fais for comments and assistance on this manuscript. Its preparation was supported by a Discovery Grant to the first author from the Natural Science and Engineering Research Council (Canada), funding from the Human Frontiers Science Program, and a Graduate Research Fellowship to the second author from the National Science Foundation (USA). We also gratefully acknowledge the support of the Canada Research Chair Program, the Canada Foundation for Innovation, and the Canadian Institutes for Advanced Research.

References

1 Fifer, W.P. and Moon, C. (2003) Prenatal development. In *An Introduction to Developmental Psychology* (Slater, A. and Bremner, G., eds), pp. 95–114, Blackwell

2 Kisilevsky, B.S. *et al*. (2003) Effects of experience on fetal voice recognition. *Psychol. Sci.* 14, 220–224

3 Tincoff, R. *et al*. (2005) The role of speech rhythm in language discrimination: further tests with a nonhuman primate. *Dev. Sci.* 8, 26–35

4 Diehl, R.L. *et al*. (2004) Speech perception. *Annu. Rev. Psychol.* 55, 149–179

5 McMurray, B. and Aslin, R.N. (2005) Infants are sensitive to within-category variation in speech perception. *Cognition* 95, B15–B26

6 Vouloumanos, A. and Werker, J.F. (2004) Tuned to the signal: the privileged status of speech for young infants. *Dev. Sci.* 7, 270–276

7 Shi, R. *et al*. (1999) Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72, B11–B21

8 Christophe, A. *et al*. (1994) Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *J. Acoust. Soc. Am.* 95, 1570–1580

9 Peña, M. *et al*. (2003) Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11702–11705

10 Dehaene-Lambertz, G. *et al*. (2002) Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015

11 Dehaene-Lambertz, G. and Gliga, T. (2004) Common neural basis for phoneme processing in infants and adults. *J. Cogn. Neurosci.* 16, 1375–1387

12 Dehaene-Lambertz, G. *et al*. (2004) Phonetic processing in a neonate with a left sylvian infarct. *Brain Lang.* 88, 26–38

13 Saffran, J.R. *et al.* (in press). The infant's auditory world: hearing, speech, and the beginnings of language. In *Handbook of Child Psychology* (6th edn) *Vol. 2: Cognition, Perception, and Language.* (Damon, W., series ed.; Kuhn, D. and Siegler, R., vol. eds), Wiley

14 Werker, J.F. and Curtin, S. (2005) PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234

15 Polka, L. and Bohn, O-S. (2003) Asymmetries in vowel perception. *Speech Commun.* 41, 221–231

16 Best, C.C. and McRoberts, G.W. (2003) Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Lang. Speech* 46, 183–216

17 Jusczyk, P.W. (2002) How infants adapt speech-processing capacities to native-language structure. *Curr. Dir. Psychol. Sci.* 11, 15–18

18 Werker, J.F. (1995) Exploring developmental changes in cross-language speech perception. In *An Invitation to Cognitive Science, Part I: Language*. (Osherson, D., series ed.; Gleitman, L. and Liberman, M., vol. eds), pp. 87–106, MIT Press

19 Pisoni, D.B. and Lively, S.E. (1995) Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In *Speech Perception and Linguistic Experience* (Strange, W., ed.), pp. 433–459, York Press

20 Rivera-Gaxiola, M. *et al*. (2005) Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev. Sci.* 8, 162–172

21 Anderson, J.L. *et al*. (2003) A statistical basis for speech sound discrimination. *Lang. Speech* 46, 155–182

22 Saffran, J.R. and Thiessen, E.D. (2003) Pattern induction by infant language learners. *Dev. Psychol.* 39, 484–494

23 Gerken, L.A. (2004) Nine-month-old infants extract structural principles required for natural language. *Cognition* 93, B89–B96

24 Chambers, K.E. *et al*. (2003) Infants learn phonotactic regularities from brief auditory experiences. *Cognition* 87, B69–B77

25 Maye, J. *et al*. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111

26 Kuhl, P.K. (2004) Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843

27 Iverson, P. *et al*. (2003) A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57

28 Kuhl, P.K. *et al*. (2003) Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9096–9101

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

30 Gerken, L. and Aslin, R.N. (2005) Thirty years of research on infant speech perception: the legacy of Peter W. Jusczyk. *Lang. Learn. Dev.* 1, 5–21

31 Swingley, D. (2005) Statistical clustering and the contents of the infant vocabulary. *Cogn. Psychol.* 50, 86–132

32 Bortfeld, H. *et al*. (2005) Mommy and me. *Psychol. Sci.* 16, 298–304

33 Thiessen, E.D. and Saffran, J.R. (2003) When cues collide: statistical and stress cues in infant word segmentation. *Dev. Psychol.* 39, 706–716

34 Curtin, S. *et al*. (2005) Stress changes the representational landscape: evidence from word segmentation. *Cognition* 96, 233–262

35 Johnson, E.K. and Jusczyk, P.W. (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567

36 Houston, D.M. and Jusczyk, P.W. (2003) Infants' long-term memory for the sound patterns of words and voices. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 1143–1154

37 Singh, L. *et al*. (2004) Preference and processing: the role of speech affect in early speech spoken word recognition. *J. Mem. Lang.* 51, 173–189

38 Vihman, M.M. *et al*. (2004) The role of accentual pattern in early lexical representation. *J. Mem. Lang.* 50, 336–353

39 Halle, P.A. and de Boysson-Bardies, B. (1996) The format of representation of recognized words in infants' early receptive lexicon. *Infant Behav. Dev.* 19, 463–481

40 Swingley, D. (2005) 11-month-olds' knowledge of how familiar words sound. *Dev. Sci.* 8, 432–443

41 Tincoff, R. and Jusczyk, P.W. (1999) Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* 10, 172–175

42 Gogate, L. *et al*. (2001) Intersensory origins of word comprehension: an ecological-dynamic systems view. *Dev. Sci.* 4, 1–37

43 Hollich, G.J. *et al.* (2000) Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monogr. Soc. Res. Child Dev.* 65, i–vi, 1–123

44 Schafer, G. and Plunkett, K. (1998) Rapid word learning by 15-month-olds under tightly-controlled conditions. *Child Dev.* 69, 309–320

45 Werker, J.F. *et al*. (1998) Acquisition of word–object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309

46 Pater, J. *et al*. (2004) The lexical acquisition of phonological contrasts. *Language* 80, 361–379

47 Stager, C.L. and Werker, J.F. (1997) Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature* 388, 381–382

48 Werker, J.F. and Fennell, C.T. (2004) From listening to sounds to listening to words: early steps in word learning. In *Weaving a Lexicon* (Hall, G. and Waxman, S., eds), pp. 79–109, MIT Press

49 Ballem, K.D. and Plunkett, K. (2005) Phonological specificity in children at 1;2. *J. Child Lang.* 32, 159–173

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

51 Swingley, D. and Aslin, R.N. (2002) Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychol. Sci.* 13, 480–484

52 Friedrich, M. and Friederici, A.D. (2005) Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport* 16, 653–656

53 Mills, D.L. *et al*. (2004) Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *J. Cogn. Neurosci.* 16, 1452–1464

54 Mills, D.L. *et al*. (1997) Language comprehension and cerebral specialization from 13 to 20 months. *Dev. Neuropsychol.* 13, 397–445

55 Xu, F. (2002) The role of language in acquiring object kind concepts in infancy. *Cognition* 85, 223–250

56 Waxman, S.R. and Lidz, J. (in press). Early word learning. In *Handbook of Child Psychology* (6th edn) *Vol. 2: Cognition, Perception, and Language.* (Damon, W., series ed.; Kuhn, D. and Siegler, R., vol. eds), Wiley

57 Beckman, M.E. and Edwards, J. (2000) The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Dev.* 71, 240–249

58 Fernald, A. *et al*. (1998) Rapid gains in speed of verbal processing by infants in the second year. *Psychol. Sci.* 9, 228–231

59 Fernald, A. *et al*. (2001) When half a word is enough: infants can recognize spoken words using partial phonetic information. *Child Dev.* 72, 1003–1015

60 *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Morgan, J.L. and Demuth, K., eds), pp. 171–184, Erlbaum

61 Carey, S. (2004) Bootstrapping and the origins of concepts. *Daedalus* 133, 59–68

62 Tsao, F-M. *et al*. (2004) Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev.* 75, 1067–1084

63 Molfese, D.L. *et al*. (2003) Discrimination of language skills at five years of age using event related potentials recorded at birth. *Dev. Neuropsychol.* 24, 541–558

64 Clarkson, R.L. *et al*. (1989) Speech perception in children with histories of recurrent otitis media. *J. Acoust. Soc. Am.* 85, 926–933

65 Houston, D.M. (2003) Development of pre-word-learning skills in infants with cochlear implants. *The Volta Review* 103, 303–326

66 Pallier, C. *et al*. (2001) The influence of native-language phonology on lexical-access: exemplar-based versus abstract lexical entries. *Psychol. Sci.* 12, 445–449

67 Pallier, C. *et al*. (1997) A limit on behavioral plasticity in speech perception. *Cognition* 64, B9–B17

68 Ventureyra, V. *et al*. (2004) The loss of first language phonetic perception in adopted Koreans. *J. Neuroling.* 17, 79–91

69 Oh, J.S. *et al*. (2003) Holding on to childhood language memory. *Cognition* 86, B53–B64

70 Knightly, L.M. *et al*. (2003) Production benefits of childhood overhearing. *J. Acoust. Soc. Am.* 114, 465–474

71 Swingley, D. and Aslin, R.N. (2000) Spoken word recognition and lexical representation in very young children. *Cognition* 76, 147–166

72 Bailey, T.M. and Plunkett, K. (2002) Phonological specificity in early words. *Cogn. Dev.* 17, 1265–1282

73 Swingley, D. *et al*. (1999) Continuous processing in word recognition at 24 months. *Cognition* 71, 73–108

74 Bosch, L. and Sebastián-Gallés, N. (2003) Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech* 46, 217–243