Cognitive Science, **1**(1):1-50 (2003)

How Does Lexical Acquisition Begin? A cognitive perspective

Chunyu Kit

Department of Chinese, Translation and Linguistics
City University of Hong Kong
Tat Chee Ave., Kowloon, Hong Kong
ctckit@cityu.edu.hk

Abstract

Lexical acquisition is a critical stage of language development, during which human infants learn a set of word forms and their association with meanings, starting from little *a priori* knowledge about words - they do not even know whether there are words in their mother tongues. How do the infants infer individual words from the continuous speech stream to which they are exposed? This paper intends to conduct a comprehensive review of contemporary studies on how the lexical acquisition starts. It first gives a brief introduction to language development, and then examines the characteristics of the speech input to lexical-learning infants and the speech perceptual abilities they have developed at the very beginning of the learning. Possible strategies of speech segmentation for word discovery and various cues that may facilitate the bootstrapping process involved in the learning, including the prosodic, allophonic, phonotactic and distributional cues, are discussed in detail, and a number of questions concerning the cue-based studies are asked: how do the infants acquire the cues for discovering words? Are the cues the starting point, or the by-product, of the learning? Is there any more fundamental cognitive mechanism that the infants exploit to induce the cues and words?

1 Introduction

Lexical acquisition is an important stage in the language development of human infants. Words are the basic building blocks for utterances. Without words, there would be no phrases, no utterances, and therefore no syntax, no semantics, and, finally, no language. Therefore, lexical acquisition is thought of as a critical initial step, if not the very first step, towards the proper development of language competence.

In general, lexical learning involves three main tasks, namely, the acquisition of word forms and meanings, and the association of the word forms with appropriate meanings. The word forms are sound sequences in speech input, e.g., string of syllables. By the term "meaning" we refer to mental representation of concepts in our mind and objects in the real world. Through a valid association of the two, speakers may use a word to refer to what they intend to. However, how meanings are represented and manipulated in our mind, the most complicated black box in the world, are still quite opaque to us for the time being. A lot of precious research focused, instead, on a very interesting and more observable part of human lexical acquisition – how preverbal infants infer word forms from speech input without any *a priori* knowledge about what words are. Although there is also no way to teach them to identify individual words in speech, they seem to have some pre-existing learning mechanism(s) to infer word forms by themselves. What amazes researchers in the field strikingly is that they seem to carry out such a difficult learning task effortlessly.

This paper is intended to review a number of critical cognitive aspects of lexical acquisition by human language learners, with highlights on the very beginning of lexical development, serving to give a psycholinguistic and cognitive background for researchers newly coming into the field or a related research area. It will first give an overview of language development as a general introduction, with an emphasis on issues closely related to lexical development. At the beginning of lexical development, two things are most critical: one is the speech input that a lexical learner is exposed to, and the other is the learner's speech perceptual ability, which determines how much the learner will receive from the speech input. The latter reveals how well a preverbal infant has prepared itself for inferring word forms from a continuous speech stream, beyond receiving the speech stream as a discrete sequence of individual sounds. Therefore, it is important to examine the characteristics of the speech input that lexical-learning infants are exposed to, and to analyse how much they can receive, by examining their perceptual ability. It is this ability that will enable them to learn to identify the first few lexical items from the speech input, starting from an empty lexicon. The input and the perceptual ability together are the basis for an infant to acquire a lexicon for understanding and producing utterances. Without adequate speech input, an infant won't be able to develop a lexicon, let alone a language; without proper speech perception, an infant cannot receive adequate speech input, and therefore cannot acquire any language properly.

In addition to reviewing existing studies on human infants' lexical acquisition, we also ask some questions. For example, it is commonly believed by many psycholinguistic researchers that a number of prosodic cues in the speech input (e.g., in child-directed speech) to the language-learning infants are the critical factors that facilitate, and perhaps even enable, their lexical acquisition. A few of our questions concerning such prosodic cues are the following. From where, and when, does an infant learn such cues? Does an infant know any such cues before knowing any words? Are these cues the starting point for lexical learning, or the result, or more precisely, the by-product, of an early period of lexical learning? If having knowledge of such cues is an indispensable starting point for lexical acquisition, a very interesting logical question would arise: without knowing words, at least some words, how could a learner know these cues are the cues for words?

Following these questions without any satisfactory answers, an even more interesting and more important question that we would like to ask is: without any cues, even the most salient word boundary markers – the pauses in between phrases and utterances – can a learner, be it an infant or a machine agent, learn a lexicon for a language from an adequate volume of speech input of the language? If we could get an affirmative answer for this question, even though the learner might not be able to get all lexical items completely correct at a time, we would arrive at a more fundamental cognitive mechanism for lexical learning at the centre of language development than prosodic cues and other constraints (e.g., phonotactics). It is highly possible that all cues for words may be derived from a more fundamental learning mechanism. What we know for sure is that these cues are not innate, but learned after birth, because different languages may have different sets of cues, and at different stages of learning a learning infant may rely on different sets of cues, e.g., when more words are learned, the learners may trust some more reliable cues than others. However, if the infants have to rely on such cues to discover new words, a more interesting question that immediately follows is how the infants acquire them for the purpose of new word discovery. By what criteria do they discriminate the cues from non-cues? What is the reliability of a given cue, and how do they handle the cases involving cues that are not 100% reliable? All these questions deserve careful studies in a cue-based approach. If lexical cues were indispensable in the explanation of how the infants learn new words, lexical cue learning would be a critical initial stage of lexical development, from which we may hope to dig out, with more research effort, of course, some more profound mechanisms of lexical learning than the direct, if not trivial, explanation that the infants learn words by some existing cues.

The rest of the paper is organised as follows. Section 2 gives a brief overview of child language development, serving to give a broad background to situate the cognitive, in particular, psycholinguistic, studies of lexical acquisition, to be discussed in later sections. Section 3 discusses a number of idiosyncratic characteristics of the speech input to lexical acquisition that might play some critical role in facilitating, triggering and even bootstrapping the very initial stage of lexical acquisition. Section 4 reviews language-learning infants' perceptual capacity for lexical learning. We are particularly interested in what abilities an infant is born with and what are later developed. Section 5 discusses some recent psycholinguistic studies on lexical acquisition about how a infant learns words with the aid of some special cues and constraints as well as inborn capabilities. Finally, we conclude the paper in Section 6, with a summary and some comments and criticisms.

2 A Brief Overview of Language Development

The fact that new-born babies of a few days old have a certain awareness of the difference between their own languages and other ones [122] and that infants a few days to a few months old are found to prefer to listen to natural language speech than other auditory inputs [33, 67] suggests that a child may begin language perception and, therefore, language acquisition, before birth. It is argued that it is the human babies' inborn sensitivity to some specific prosodic properties in natural language speech that enables them to discriminate speech in their mother tongues from speech in other languages [122]. It is also noted that at the beginning infants are sensible to all sound contrasts in all natural languages, but later this ability fades and the infants' speech perception gradually adapts to the phonology of their mother tongue [63].

Infants are born with a nascent structure-seeking mechanism to discover particular units with particular distributional patterns in natural language input, guided by innately specified structural constraints [153, 148, 82, 149, 150, 154, 151]. This mechanism is sensitive to the patterned organisation (e.g., rhythmic, temporal and hierarchical organisation) of natural language phonology common to all languages [60], be they spoken or signed [151], and the linguistic units and patterns learned by such mechanism correspond in size and organisation to phonetic and syllabic units common to all languages [151]. This direction of exploration is seen to follow the *innateness hypothesis* in Chomskyan linguistics that human species have an innate genetic endowment, known as *Universal Grammar* (UG), for acquiring natural languages [26, 27, 28, 29, 30].

A normal child starts to produce reduplicative *babbling*, composed of repeated syllables (*bababa, dadada*, etc.), within 6 to 10 months after birth. Some intonation patterns and some imitation of adults' speech are observed to appear during the late babbling stage, from 9 to 12 months. Interestingly, deaf children also begin to babble with their sign-language at a similar age, producing sequences of syllabic units[1] that are observed to be fundamentally identical to vocal babbling produced by normal children who have an ordinary exposure to a spoken language [152]. Based on the observation of no significant difference between the acquisition of spoken and signed languages, it is argued, or suggested, that human infants' capacity for language acquisition is not specific for speech, rather, it is part of the children's cognitive capacity for acquiring abstract structures, including linguistic structure, from the surrounding world [138, 150, 152].

Children can understand *many* words long before they produce the *first* word [79]. Usually, a child produces *one-word* utterances roughly by the age of 10 to 11 months. The gap between comprehension and production is lexically great at this stage: a child may be able to understand about one hundred words when it starts to produce the first word [4]. The majority of the children's early lexical items are names of individuals and objects in their environment, such as

[1] "As in spoken languages, signed languages are constructed from a finite set of meaningless units (phonetic units); $\cdots$ ASL's phonetic inventory is drawn from the four parameters of a sign – handshape, movement, location, and palm orientation – each of which contains a restricted set of phonetic units (for example, a set of handshapes, a set of movements). Phonetic units are further organised into structured units called syllables." "A well-formed syllable has a handshape, a location and a path movement (change in location) or secondary movement (change in handshape or orientation)" [152] (pp.1495, Note 20). ASL is American Signed Language, and "a sign [in signed languages] has identical linguistic properties to a word in spoken languages" (*Ibid*, pp.1494).

“Mama” and “car”. Some action verbs, which refer to actions that frequently take place around the children or related to them (e.g., “give”, “hit”, “drink” and “eat”), and a few adjectives (e.g., “big” and “good”) are also in the lexicon. Abstract words come into the lexicon much later. At the age of 16-18 months, the single word utterances seem to show some semantic categories (e.g., agent, action and object) [80], but have a very vague mapping to adult meaning, for example, a short utterance “cup” may mean “a cup is there”, “I see a cup” or “I want the cup”.

By the age of about 18 months, children start to produce *two-word* utterances, or more precisely, two-word phrases. The children also undergo a “word-spurt” or “naming explosion” at this stage. The number of words in a child’s lexicon increases rapidly in this period of language development. There is also evidence that the emergence of phrases in the child speech correlates to the word-spurt [135, 3]. A correlation between word learning and initial syntactic development is observed. For example, young children are sensitive to syntactic information (e.g., part of speech) when inferring a new word’s meaning. They tend to guess that a verb-like new word refers to an action, a countable noun to a physical object or an individual, and a mass noun to a kind of substance or a piece of non-individual entity [19, 103, 165].

Next, children start to produce *telegraphic speech* [21]. The term “telegraphic speech” is specifically used in the study of child language development to refer to children’s short utterances that lack grammatical inflections and functional (or closed-class) words and/or morphemes, like determiners (e.g., “the”), prepositions (e.g., “of”) and suffixes (e.g., “-s” and “-ed”) [22]. Such utterances sound like telegrams using as few words as possible to convey essential meanings. Children tend to use telegraphic forms even when they are trying to imitate adults’ full sentences. Omission of subject is a frequent phenomenon in children’s telegraphic speech [10]. Closed-class morphemes are observed to have a relatively fixed order to show up in children’s speech, e.g., in English, the morpheme “-ing” (as in “(is) talking”) for present progressive shows up earlier than third-person singular “-s” (as in “talks”) [20, 48]. The order (or tendency) of acquisition of a few frequent grammatical morphemes in English is: the present progressive “-ing” appears first, then the regular plural “-s”, possessive “-s”, irregular past tense forms and regular forms [178]. It is noted that semantic complexity and phonological salience have a certain influence on the emergence order of these morphemes [49].

A very important phenomenon observed at this initial stage of grammatical competence is that children rarely make mistakes about *word order* [14, 13, 9, 20, 153, 11], indicating that children start to understand and master some fundamental syntactic properties in adults’ speech. There is evidence to support the idea that children have a certain grasp of word order knowledge even prior to their producing of telegraphic speech [74]. More interestingly, children even attempt to utilise word order to express grammatical relations at the initial phase of acquiring “free word order” languages, where grammatical relations are marked by case markers [138, 153].

Children start to produce multiple word utterances by the age of about two and half. The average utterance length goes up gradually in the next few years. Accordingly, more and more functional words are used, and the utterances produced by children become more and more complex, in terms of grammatical structure. Yes-no questions, relative clauses and control structures (e.g., "I want him to come") appear in the children's speech. Children of five and six years old are able to add appropriate grammatical suffixes to words according to their grammatical classes, which can be induced from the context. This is known as the famous *wug procedure*, where *wug* is a word invented for experimental purposes [5]. Children also attempt to use, and invent, rules to infer morphological forms for a verb according to time and tense. Interestingly, however, they make morphological mistakes, known as *overgeneralization.* Some frequently quoted examples are "goed" (*vs.* "went"), "comed" (*vs.* "came") and "mans" (*vs.* "men"). Inappropriate usage of words also occurs in child speech, e.g., "I giggled the baby" instead of "I made the baby giggle". Many parents try to correct their children's mistakes like these. However, there is a consensus in child language development that parents' correction of children's speech mistakes, known as *negative evidence*, does not have any essential influence on the children's language acquisition [72, 130, 12, 71, 116]. It is very interesting that in some cultures, adults do not talk to children before they have a full competence to speak, let alone correct their speech mistakes.

After word spurt, children's vocabulary grows steadily. It has been estimated that a child acquires about nine words per day from the age of 1.5 to 6 years [24]. It is also observed that there is a critical period for human language development. If a child has no chance to be exposed to any language by the onset of adolescence, s/he seems to have a very slim chance to retain the ability to learn to speak in a language with full-fledged syntax – there is empirical evidence for this from a number of cases of feral or isolated children, e.g., the Wild Boy of Aveyron [104] and Genie [36, 158]. If exposed to a language (including sign language) after the age of 7, children have less and less chance to become totally fluent in native accent. It is reported that few Chinese and Korean children who immigrated to USA after the age of 7 can become totally competent in American English [81]. A similar result is found with people acquiring ASL as their first language [136].

3 Speech Input for Lexical Acquisition

Infants who are born with the ability to learn natural language will not really learn a language if they are not exposed adequately to natural language speech. Two factors are critical in this exposure to enable the learning: one is speech input to the learners, and the other is what the learners really receive from the input based on their speech perception capabilities. In this section we will analyse the characteristics of speech input to language-learning infants at the early stage of language development, in particular, the input for lexical acquisition, and discuss what influence such input has upon the infants' learning. In the next section we will review the pre-linguistic infants' speech perception. The speech perception determines what an infant really receives from the surrounding speech environment.

From the time of their birth, preverbal infants are surrounded by adults' speech, directed to other adults or to the infants. Young infants are reported to prefer to listen to infant-oriented "motherese" than to normal adult-to-adult speech [55]. The term *motherese* is used to refer to a kind of slow high-pitched speech with smooth, exaggerated intonation contours that adults tend to use to communicate with infants in many cultures [137]. It is argued in [56] that there exist certain universal speech patterns across languages and cultures in adults' speech to infants, to soothe or express praise or disapproval with different prosodic contours – this kind of speech was even considered as a universal signal system that was believed to be based on human biology, a system of calls independent of the meaning carried by the speech but having a certain effect on children (including directing their attention or calming or arousing them).

Motherese, also known as *baby talk* and *infant-* or *child-oriented speech*, is characterised by many special features (as discussed in [62, 96, 167, 59, 60] and many others), e.g.,

- Slower speaking rate
- Higher pitch
- A wider range of fundamental frequency
- Highly varied intonations, with significant exaggeration
- More frequent and longer pauses
- Simpler and shorter utterances
- More repetitive
- Special baby-words (e.g., *doggy* and *pussy*)
- More frequent onomatopoeia and interjections
- Restriction of topics to those relevant to a child's world

The role of infant-directed speech in language acquisition, in particular lexical acquisition, is highly arguable. There is experimental evidence that young infants have a preference for listening to infant-directed speech over adult-directed speech. An experiment is reported in [55] that the 4-month-old infants in the experiment chose, by turning their heads, to listen to an infant-directed speech tape more frequently than to listen to an adult-directed speech tape. A further study in [57] found that this preference is strongly associated with the melody in the infant-directed speech, because the infants' preference was observed to persist even when all but the melody was filtered out of the speech signal. It was also found later that even in the first month after birth, the new-born babies also had the preference for infant-directed speech [34], but to these younger infants, prosody alone appeared insufficient to maintain the preference – it was observed to be only associated with the full speech signal [35]. The findings in [57] and [35] suggest that new born babies receive the infant-directed speech signal as a whole at the very beginning, but by the age of about 4 months, they have learned, somehow, to isolate the pitch contours from the whole speech signal and have built up some kind of correlation between these contours and the positive interactions with their mothers – it is this correlation that leads to the preference for the prosodic contours in the infant-directed speech.

There are many conjectures about the role of infant-directed speech (in particular, its prosodic features) in the early stage of language acquisition, based on the young infants' preference for, and sensitivity to, the prosodic features in infant-directed speech. It is said that the correlation between various types of intonation and their effects may provide the first basis for children to understand the sound-meaning correspondence. It is also proposed that the significantly exaggerated intonation and stress patterns may provide important clues to help, or even trigger, the infants' identification of lexical items and linguistic structures such as phrases and sentences in speech. The hypothesis that language learning heavily relies on the prosodic characteristics of the speech input to the learners is known as the *prosodic bootstrapping hypothesis* [66, 147, 126, 75, 65, 58] and has received considerable attention in recent years.

Many psycholinguistic experimental results were interpreted in a way to support this popular hypothesis. For example, the experimental results presented in [75] indicate that 7- to 10-month-old infants prefer to listen to motherese utterances not interrupted by a pause over the other utterances interrupted by a pause, were interpreted as an indication that clauses are salient perceptual units for young infants. A subsequent study [97], using the same procedure as in [75], further found that preverbal infants only preferred the uninterrupted utterances over the interrupted ones in motherese, but not in adult-directed speech. This result is interpreted as suggesting that pre-linguistic infants are able to identify the clauses in motherese but not in adult-directed speech [76]. If we accept these two interpretations, we might conclude that without motherese as input, infants would not be able to detect clause boundaries and therefore could not learn a language.

However, there is strong evidence against this critical role that motherese would play in language acquisition: it is reported that in a number of cultures, for example, in Papua New Guinea [160, 161], Samoa [143, 162] and even among some African Americans [73], infant-directed speech is not available to pre-linguistic infants because adults simply do not talk to the infants before they have learned to speak. The over-estimation of the importance of infant-directed speech excludes the possibility that human babies in a culture with little motherese available still can learn a language purely from the adult-directed speech taking place around them.

What we can conclude, in consideration of the positive and negative evidence discussed above, about the role that infant-directed speech may play in language acquisition is as follows. First of all, there is evidence that infant-directed speech, if available, does facilitate language acquisition at an early stage. The prosodic packaging of some salient speech units like utterances and phrases appears to have a beneficial effect on leading the learning infants to realise the existence of such linguistic structures in fluent speech. However, no matter how important the infant-directed speech may be to some stage of language development, we still do not have any grounds to state that language acquisition has to rest on the availability of speech input of this special style. Although speech of this style does take place in many cultures and does appear to have beneficial effect on the early stage of language acquisition, the fact that infants in some other cultures with little infant-directed speech available do succeed in learning their language from adult-directed speech indicates that their language faculties have enough capability to deal with natural language speech input of any style or characteristics for the purpose of acquiring a language.

Moreover, we have not had any clear idea about what causes an infant to prefer to listen to infant-directed speech. Is it an innate bias, or a preference due to prenatal exposure to human speech? If it is an innate bias, a new-born baby should equally prefer baby talk in any language. Why does it prefer only baby talk in its own mother tongue? If this preference is not an innate bias, it must be something that has to do with prenatal speech experience, that is, it is a result of learning (or memorising). Also, we still do not know what effect on later language acquisition would result from lacking such a listening preference. Does an infant have any acquisition barriers or special difficulties in any later stage of language acquisition if it is given no chance to build up such a preference? There seems to be little evidence supporting an affirmative answer to this question.

Furthermore, even in the speech environment with infant-directed speech available, it is still unlikely that language-learning infants hear only infant-directed speech and no adult-directed speech – notice that there is so much adult-directed speech surrounding them spoken by their parents and other elders. We still need to obtain a clear idea of how much an infant learns from infant-directed speech and how much from adult-directed speech. What is very likely is that infants hear a greater volume of adult-directed speech than infant-directed speech, because adults speak faster to each other and the time they speak in the baby talk style is relatively short. More importantly, when language-learning infants can talk, they do not talk first in the infant-directed style and then move to the adult-directed style. Rather, they seem to straightforwardly go into the adult-directed style: even when they are producing baby-words, they tend to use a normal prosody rather than an exaggerated one, with almost no onomatopoeia and interjections (e.g., *oh* and *uh*) – although they always hear speech that has exaggerated prosodic characteristics and is full of onomatopoeia and interjections. Also, the infants learn many lexical items and syntactic patterns that seldom or even never appear in infant-directed speech. These observations indicate that the infants are quite clear in their minds that the target language for them to learn is the adults' speech, not the baby talk.

It is argued in [38], based on the observation that many of those characteristics of child-directed speech, including the high frequency of phonological elisions and assimilations, are exactly the characteristics of adult-directed spontaneous speech in contrast to rehearsed speech heard on the radio and television, read speech in news broadcast and in lectures, that the infant-directed spontaneous speech should be considered to lie on a general continuum at the same end with the adult-directed spontaneous speech, with the former in a more extreme position. Although the phonological elisions and assimilations in spontaneous speech make the segmentation problem even harder for children, it is far from true that infants can learn better from rehearsed or read speech than from spontaneous speech.

In short, infant-directed speech is no doubt a beneficial input to language-learning infants in many cultures. In particular, its prosodic characteristics such as the exaggerated pitch contours (e.g., pitch declinations) and lengthy duration of vowels at utterance ends and (prosodic) phrase ends appear to play the role of signalling the boundaries of some salient linguistic structures to the learning infants. However, infant-directed speech is not the only input to language acquisition. We should not neglect the role that adult-directed speech may play in language acquisition: Even without infant-directed speech, human infants have no problem acquiring a language from adult-directed speech around them. Thus, a better understanding is that both infant-directed and adult-directed speech are at the same end of spontaneous speech on a general continuum and that language acquisition, including lexical learning, does not entirely rely on the availability of infant-directed speech. It is reasonable that, ultimately, children learn more from normal adult-directed speech than from special infant-directed speech, although the latter seems to play a more critical role in bootstrapping the initial stage of language acquisition.

4 Pre-linguistic Infants' Speech Perception

Human infants are born with many amazing perceptual abilities that enable and facilitate their language learning. It is argued that infants are born with a nascent structure-seeking mechanism to discover particular sized units with particular distributional patterns in the input, be it spoken or signed, and this seeking procedure is believed to be guided by some innately specified structural constraints [151] (e.g., children's early lexical use is already constrained along the bounds of word types, like object names, property words, events words). When an infant is born, this nascent mechanism is sensitive to phonological patterns (rhythmic, temporal and hierarchical) that are common in all languages [60], and is particularly sensitive to the syllable-like structures (in terms of their size and distributional patterns) in the input (spoken or signed) [152].

However, this nascent structure-seeking mechanism for language learning would not work if the atomic elements, i.e., speech sounds, in the speech input could not be correctly detected by pre-linguistic infants. While thinking about how infants start to acquire lexical items from speech input without any supervision, the first question we have to ask is how strong is the pre-linguistic infants' ability to discriminate among speech sounds in terms of their phonological characteristics (or features) when they start to learn a language.

Just as humans are born to see, they are born to hear. An infant's inner ear is physiologically fully grown at birth. It is reasonable to assume that infants can hear even in the womb. There is evidence that even before birth the foetus' auditory system is functioning. It is observed that the foetus in the womb responds to external sounds [101], and also that newborns show preference in their first day for their mother's voice over an unfamiliar female's voice [50]. This preference cannot be explained if the newborns have not had any auditory experience. Two stronger pieces of evidence supporting the idea that infants have started speech perception, and therefore language learning, before their birth are the demonstrations that the newborns can discriminate a passage loudly read by their mothers during the last six weeks of their pregnancy from an unfamiliar passage, even when it is read by a woman other than their mothers [52], and that they can distinguish utterances in their mothers' language (e.g., French) from utterances in a foreign language (e.g., Russian) [122]. The prosodic contours in their mothers' speech are observed to play an important role in enabling the infants to recognise utterances in their mother tongues. It is reported in [122] that even when the speech samples were filtered in a way to only keep the prosodic (in particular, rhythmic) information, the results turned out to be similar.

This listening preference must be shaped by prenatal speech exposure in the uterus. There are intra-uterine recordings to reveal the fact that the low-frequency components of sounds from the outside world, in particular the maternal speech with its distinctive rhythmic features, are audible in the uterus. It is found that late-term foetuses, in the last three months of gestation, respond to external speech stimuli in a rather consistent way [106, 105, 108]. It is reported in [107] that repeatedly presenting a pair of French syllables ([ba] and [bi], or [bi] and [ba]), uttered by a female, to foetuses of 36-40 weeks every 3.5 seconds elicited a deceleration of heart rate, and when the order of the two syllables was reversed, the same deceleration reliably showed up after 16 presentations. This fact was interpreted as an indication that the foetuses could discriminate between the two types of stimulus. Near-term foetus' heart rate responses to acoustic stimulation were further studied using the short sentence "Dick a du bon thé" ("Dick has some good tea") in a male and a female voice [108]. It was observed that 77% and 66% of the subjects reacted, respectively, to the male and female voices with a declarative heart rate change within the first 10 seconds of the stimulation. When the initial voice changed, 69% of the subjects showed a heart rate deceleration, whereas 43% of the control subjects, who kept hearing the initial voice, showed slight acceleration [109]. These findings indicate that the near-term foetuses can differentiate between the two voices and that the foetal auditory system has matured to such a degree that it can detect an acoustic change in human speech based on a rather small speech sample. It is also reported in [51] that mothers' loud recitation of one of two selected rhymes to their 37-week old foetuses once a day for a period of four weeks could result in the familiar rhyme causing the foetuses' heart rate to decrease consistently when they were listening to the familiar rhyme, whereas the unfamiliar rhyme did not have such an effect.

Infants' perceptual ability develops rapidly after birth. Within a few weeks of birth, they are able to make distinctions between human voice and other sounds, and in about two months, they can detect the difference between angry and friendly voice qualities [113]. Experiments using the *non-nutritive sucking* technique (also known as the *high-amplitude sucking* (HAS) technique, as in [76]), and the *head-turning* technique in the past three decades have greatly deepened our understanding of pre-linguistic infants', in particular, newborns', perceptual abilities related to speech. These techniques show that in the first month infants can distinguish voiced from unvoiced sounds [54], and can discriminate vowel contrasts (e.g., /i/ *vs.* /a/) [168] and consonant contrasts (e.g., /p/ *vs.* /b/) [100]. In two months they are able to tell apart different intonation contours and places of articulation [131, 32].

Initially, infants' speech discrimination abilities are language-general, rather than specific to any language. The infants can discriminate contrasts outside the ambient language. It is reported that English babies can discriminate consonant contrasts that exist in Hindi [175] and vowel contrasts that exist in French [169], but not in English. A comprehensive review of pre-linguistic infants' perceptual abilities for sound contrasts from many languages is given in [68] and in [84].

However, although the pre-linguistic infants' hearing is so sharp that speech sound contrasts in any language can be detected, they do not perceive speech sounds as distinct individual sounds. Rather, they perform categorial perceptions while perceiving speech sounds: speech sounds with acoustic differences within a certain range are recognised as a *phoneme*. From the acoustic perspective, human speech signals are sound waves transferred through air pressure changes from a speaker's vocal organs to a listener's ears. Human ears – more precisely, human brains, which receive speech signals from the ears – do not interpret sound waves of human speech as waves, but as *segments* or *units* that bear phonological significance in a language. These basic units, or sounds, are known as *phones* or *allophones*. In human speech perception, phones are grouped into phonemes that are phonologically distinct from each other. In this sense, humans as natural language speakers only "hear", or perceive, phonemes, rather than phones, in general. Every phone must be heard as a phonemic category. People cannot distinguish sounds with a 20-msec difference of *voice onset time* (VOT) within the same phonemic category, e.g., variant instances of /p/, but can detect a 20-msec difference across a *phoneme boundary* – for example, as demonstrated in [177], a sound in the /b/ and /p/ categories (whose phoneme boundary is at about 25 msec VOT) will be heard as a /b/ if its VOT is shorter than the boundary point (i.e., somewhere near 25 msec), otherwise, it will be heard as a /p/.

It is showed that infants' categorial speech perception is similar to that of adults [54], although their phonology may not be identical to that of the adults. The more an infant's speech perception has been attuned to the adults' phonology, the less sensitive is its hearing to the sound contrasts not existing in its mother tongue. It is showed in [176] that English-learning infants 6 to 8 months old could discriminate consonant contrasts that exist in Hindi and Inslekepmx (a language spoken by the native Salish of British Columbia) but not in English, but very few English-learning infants of 10 to 12 months old could do so.

This categorial perception is not unique to speech. Some non-speech sounds, such as noise-buzz sequences, are also found to be perceived categorially [124]. And, the categorial perception of sounds is not specific to humans; it is reported that chinchillas, a kind of rodent, can also sense the phoneme boundary effect between /p/ and /b/ [102]. However, although the chinchilla's aural system is observed to function in a similar way as human's, chinchillas (and many other mammal species which have a similar mechanism for aural perception) do not develop any spoken language relying on this categorial perception mechanism. Thus, it is concluded that the categorial perception of sounds, as reflected in the phoneme boundary effect, is a property of the mammalian aural system that can be utilised to develop languages as signal systems for communication, rather than a unique linguistic property of human's auditory perception [101, 125, 76]. Nevertheless, such perception reflects the extent to which the human species has prepared, through evolution, for language acquisition. Without this particular endowment, humans would not have had any spoken language like those in use today, let alone language acquisition.

Although humans are born to have, in general, the ability to discriminate phonetic contrasts in any language (no matter how subtle these contrasts are) and the ability to perform categorial perception of speech sounds (in particular, the consonants), they do not seem to perceive speech phone by phone, but syllable by syllable. There is evidence that syllables are the basic units of mental representation of speech sounds and are, therefore, the effective units in the young infants' speech perception. It is showed in [89] that, while presenting randomly ordered sequences of syllables to 2-month-olds, the infants increase their response to new syllables, no matter how subtle the change is from the old syllables to the new ones, but the change of the number of the phonemes does not increase the response. A perhaps even stronger piece of evidence supporting the view that syllables are young infants' effective mental representations for speech sounds is the study [8] on how 4-day-old infants categorise multi-syllabic utterances: the babies had no difficulty distinguishing sound sequences of different numbers of syllables, e.g., two syllables (e.g., *rifo*) and three syllables (e.g., *kesopa*), but could not detect the difference between two-syllable sequences with four phonemes (e.g., *rifo*) and two-syllable sequences with six phonemes (e.g., *treklu*). These studies suggest that, although young infants are highly sensitive to phonetic contrasts and to phoneme boundaries, they tend to, interestingly, nevertheless, perceive a speech stream syllable by syllable, instead of phoneme by phoneme.

Since syllables are found to be the effective units in speech representation, a speech stream can be understood as a sequence of individual syllables. Naturally-occurring utterances produced by adults are sequences of continuing syllables, rather than discrete sequences of isolated syllables. Therefore, prelinguistic infants' ability to identify individual syllables in a steam of on-going speech is critical to their language development, in particular, in the initial stage. An experiment in [70] using the head turning technique showed that prelinguistic infants of 6.5 months old can discriminate individual syllables like [ba] and [du] not only in isolation but also in combination with other syllables. The experiment also revealed that, when the target syllables were embedded in more complex contexts, the discrimination of them appeared more difficult. For example, the correct rate of the infants' discrimination of [kokodu] from [kokoba], where the target syllables are embedded in a redundant sequence, is 75%, but the correct rate for the recognition of target syllables in mixed sequences such as [kotiba] and [kotidu] goes down to 67%. Another study [95] shows that prosodic features, such as stress on target syllables, e.g., [koti*ba*] and [koti*du*], can help young infants to do the discrimination better than on the same multi-syllabic sequences with an even intonation.

However, the recognition of syllables as the basic units in infants' speech representation does not indicate that there is absolutely no segmental representation for any speech sounds. It seems necessary to have some segmental representation in many languages. For example, in English, we have a few morphemes of certain syntactic importance whose sounds do not constitute a syllable due to the lack of a vowel, e.g., *'s* as the abbreviated form for the third person singular copula *is* (as in *it's here* and *that's it*), *-s* as the suffix for plural nouns (as in *trees*) and *-ed* as the suffix for past tense verbs when not preceded by a vowel (as in *trained*). It is reasonable that morphemes of this kind in a language may need to have a segmental rather than a syllabic representation. From this perspective, a speech stream can be represented as a continuous sequence of phonemes, and a syllable in the speech can, accordingly, be thought of as a number of phonemes in a structure, e.g., in a CVC structure.

In summary, human infants are born with a remarkable sensitivity to sound contrasts existing in natural language speech and with a special ability to perform categorial perception of speech sounds, although they are born with very little speech experience, if not completely none. A language-learning infant's speech perception is gradually attuned to the phonology in its ambient language as it is exposed to more and more speech data. In addition, there is evidence that syllables, instead of phones, are the representational and perceptual units perceived in the speech stream, i.e., syllables are the basic chunks of sound that the infants perceive, although this observation does not deny the fact that young infants are able to distinguish phonemes within syllables and may also receive a syllable as a structure of phonemes. This evidence also explains why so many prosodic features over syllables, such as stress and intonation contours, play a critical role in initiating and facilitating the pre-linguistic infants' lexical acquisition. In next section we will focus on the prosodic and other cues that facilitate speech segmentation by the infants for the purpose of lexical acquisition.

5 Speech Segmentation and Word Discovery

Lexical acquisition involves at least two aspects; one is the identification (or determination) of word forms in spoken utterances and the other is the appropriate association of lexical properties (e.g., meaning or conceptual information, syntactic properties like part-of-speech and agreement) with each word form.

There is a *bootstrapping* problem in infants' acquisition of lexical forms. It is reasonable to assume that they have an empty lexicon at the beginning of language acquisition. However, the question is: how do they get the learning task started? In order to acquire words one after another to build up a lexicon, the infants must have the ability, without knowing any particular words at the very beginning, to recognise word forms in the speech stream. To our knowledge, one must know the word forms before one can recognise the words in the speech. So the problem is: how can infants, who know no words, extract words from the speech to which they are exposed, and put them in the lexicon one by one so as to enlarge lexicon day after day? There seems to be a chicken-and-egg problem for preverbal infants to resolve: they must be able to segment fluent speech into words in order to develop a lexicon from scratch, but they must know a certain number of words before they can perform the segmentation! How can an infant get around this problem and develop its lexicon? That is why lexical acquisition is so interesting and perplexing. It is not as trivial a problem as it seems to be – just pick out words from the speech input and put them into the lexicon. It is not as simple as this. It involves a deadlock problem: if you know no words, you can't pick any words from the speech input; if you can't pick any words, you remain knowing no words.

It seems unlikely that we can find out whether the chicken or the egg appears first. A possible way out of this dilemma is to assume that the infants have some means or strategies (e.g., the utterance-as-word strategy) to segment speech into words or word-like units for the purpose of developing their own lexicon, with the aid of various kinds of cues in the speech stream, e.g., pauses, stresses, that they are born sensitive to or that they can somehow learn to detect after birth. This early lexicon may be different from the adults' at the infants' early age; for example, some collocations of frequently co-occurring words may be taken as individual words in a child's lexicon. But this lexicon will gradually converge to the adults' lexicon along with the growth of the infants' language experience and competence. This convergence is a very interesting process of learning that is worth more exploration in the field of language acquisition.

In this section, we will take a close look into what cues in adults' speech can facilitate infants' lexical learning, and what strategies the infants can exploit to develop their lexicon into one as close to the adults' lexicon as possible. We start with the speech segmentation problem in the next subsection.

5.1 Speech Segmentation

Understanding spoken utterances involves a process of identifying discrete words from the speech stream. Only after individual words in an utterance are properly recognised can the structure of the utterance be analysed and its meaning be interpreted. Although it seems effortless for adult listeners to carry out the word recognition task during speech comprehension, it is by no means a trivial task. In addition to having to cope with troubles caused by some undesirable characteristics of the speaker's voice (such as dynamically variant speaking rate, accent, co-articulation of adjacent words, etc.) and background (speech) noise while listening to the speech signals (which can be understood as a sequence of phonemes, or syllables), the listener also has to map this sequence of continuous speech signals onto a sequence of lexical items from the listener's own lexicon, which is usually of tens of thousands of words. That is, the listener has to segment the speech input into fragments such that each fragment matches an existing word in the lexicon. The situation could be more complicated if the speech input involves any unknown new word(s).

If formulated as a hypothesis selection problem in terms of some objective function $f(\cdots)$, the speech segmentation problem can be thought of as selecting a sequence of words from the lexicon to cover exactly the input speech signal S such that the objective function on the words can be maximised. It can be expressed by the equation below:

$$Ws(S) = \underset{w_1 \circ \cdots \circ w_n = S}{\arg\max} \; f(w_1, \cdots, w_n) \tag{1}$$

where $Ws(S)$ denotes the resulting sequence of words from the input S and $\circ$ is a concatenation operation. If the object function is to evaluate the probability of the word sequence, it can be rewritten as (2), following Bayes rule.

$$f(w_1, \cdots, w_n) = \prod_{i=1}^{n} p(w_i | w_1 \cdots w_{i-1}) \tag{2}$$

To compute this objective function, a probability distribution over words given a preceding context in a language must be given or estimated somehow, for example, based on individual word sequences' relative frequencies.

Notice, however, that a constraint on on-line lexical processing for speech comprehension that is not taken into consideration in (1) is that the determination of individual words is to be done one by one in order: once a word is determined, the listener moves on to work on the next word. In general, backtracking to any previous word is not permissible in real time speech processing by human listeners.

The difficulties in speech segmentation and word recognition lie in the fact that word boundaries are not explicitly marked in the speech signal: not only are there no explicit markers (e.g., pauses) about where a word begins and ends, there are also no cues that are fully reliable, as noted in [111, 133] – although there exist many types of cues in continuous speech to facilitate the location of word boundaries, there are always many exceptional cases where the cues do not work right. Useful cues include lengthening of word-initial and -final syllables, allophonic cues (e.g., aspiration of word-initial stop consonants), phonotactic cues – disallowable sequences (esp., pairs) of consonants in words (or syllables) (e.g., [mr] in English), and many others. Importantly, although languages differ from each other also in terms of their rhythm and prosody, the particular metrical structure of a language can be made use of to facilitate the segmentation. For example, the syllable is the basic metrical unit in French and Spanish, and the mora is the basic metrical unit in Japanese [145, 45].

There is evidence that the native speakers of these languages make use of the syllabic and moraic information, respectively, to perform speech segmentation [42, 43, 146, 46].

In English and Dutch, the distinctive rhythmic characteristics of strong and weak syllables are utilised by native speakers to do speech segmentation [44, 39, 38, 121, 140, 173, 174]. There are observations that more than 90% of content words in English start with a strong syllable, about 75% of the strong syllables are at word onsets in English speech [40] and that about 85% of Dutch words have a strong syllable at the word onset [163, 173]. It is reasonable to infer, based on these findings, that the native speakers of these languages tend to speculate a word onset at a strong syllable. Accordingly, the *metrical segmentation strategy* (MSS) is formulated to characterise the native speakers' bias in pre-lexical processing of speech input [44]. There is evidence from a number of investigations by Cutler and her co-workers that English adults do apply the MSS to predict word onsets with the occurrence of strong (or stressed) syllables in speech processing [37, 39, 38].

In addition to the tendency of using the MSS, adults' segmentation of fluent speech is also full of other activities such as activation of candidate words at all points along the speech input and competition among activated candidates. For example, when *can* is heard, many words beginning with the syllable *can*, such as *can*, *cancel*, *candle*, *canteen*, etc., are activated. When more speech signals are received, some candidates will be ruled out if they are inconsistent with the new signals, and some continue to survive until the end of the utterance. The competition takes place not only among individual candidate hypotheses activated at the same point, but, more importantly, also among difference parses, each being a sequence of activated words, over the same speech fragment (e.g., a phrase) or the entire utterance. For example, when a speech fragment like *met a fourth time* is heard with some background noise (which may cause *th* to be confused with *f*), theoretically, another possible parse over this fragment could be *metaphor f time* (borrowed from [141]). If word embedding is also considered, the situation can be more complicated. How do human listeners resolve this kind of ambiguity in pre-lexical processing? A constraint called the *possible word constraint* (PWC) is proposed in [141] to model human subjects' decision making in such cases: impossible words are disfavoured. That is, a parse with all chunks being possible words in the listener's lexicon is preferred over a parse with some chunks being impossible (or unknown) words. As far as the preceding example is concerned, human subjects will tend to follow the PWC to choose *met a fourth time*, because in the other choice the chunk *f* is an impossible word in English. Also, evidence is given in [141] that it is easier, as shown by response time and correct rate in experiments, for human subjects to segment *vuffapple* into *vuff apple* than *fapple* into *f apple*, because *vuff*, which contains a vowel, is possibly a word in English whereas *f* is known for sure not to be a word – it is common sense that every word in English has a vowel. Even worse, an *f* standing alone does not even make a syllable, let alone a word.

There are a few computational models to simulate adult speech segmentation, for example, TRACE [119] and the Shortlist [142], implemented in neural networks with an emphasis on modelling the competition between candidate words. The Shortlist was extended later in [141] to incorporate the PWC. Here, we are not going into the details of these models beyond the scope of our research on lexical learning.

As shown in the brief review above, speech segmentation and lexical recognition performed on continuous speech by adult listeners, who can be thought of as equipped with a huge (if not almost complete) lexicon, is by no means simple or trivial. Many useful cues, none of which are entirely reliable though, are utilised, and a number of complicated cognitive processes, e.g., candidate word activation and competition, are involved. The listeners have some language-specific strategies, e.g., the MSS for English and Dutch, to facilitate the processing. Their cognitive behaviours in pre-lexical processing also appear to observe certain constraints, e.g., the PWC.

With regard to the complexity in adult speech segmentation, there is reason to believe that speech segmentation and word discovery by pre-linguistic infants, who have an empty lexicon, is even more difficult and complicated. How do they come to know there are words in their language? What cues can they make use of to discover words in fluent spontaneous speech? Are there any specific strategies that they can exploit?

5.2 Cues in Speech for Word Discovery

Although how young infants acquire word forms from a continuous speech stream is recognised as a central task in lexical acquisition, it is not yet clear so far in psycholinguistic studies how and when the infants start to be aware of the existence of words in their language and then attempt to segment fluent speech into individual words. What we are clear about are, at least, the following facts: first, all languages have words, of which individual utterances are composed – thus, to understand an utterance one has to decompose the utterance into words and then retrieve each word's meaning from a mental lexicon; second, adult speakers do not, and also appear unable to, explicitly tell the infant language learners which sound sequences are words, in particular, where a word starts and ends, even in the situation of teaching them new words.

There seem to be two possible ways to inform lexical-learning infants of word boundaries: the first is to tell them, implicitly, by speaking in isolated words, isolated by significantly long pauses of silence; the second is to tell them explicitly by speech. Across all languages in the world, adult speakers do not speak in isolated words, even in the situation of speaking to infants in the infant-directed speech style. If they were to speak in isolated words, whether their offspring could learn their language would be seriously in doubt, because such speech loses many prosodic, rhythmic and other characteristics of frequency-and-volume change (e.g., pitch contours and intonations) in the continuous speech of their language that are known to be very useful in bootstrapping the infants' sensitivity to speech units such as clauses and phrases in the language – these units bear special prosodic demarcations to which the very young infants are sensitive [84]. If one attempts to directly tell the infants about word boundaries, one would have to speak to them in speech and they would have to segment the speech into words for understanding – in order to achieve this, however, they have to learn to do speech segmentation first. Thus, a deadlock problem arises. It appears that there is no effective way to teach preverbal infants to do speech segmentation for the purpose of lexical learning. They have to get around the deadlock problem somehow by themselves.

The greatest difference in speech segmentation between adults and the very young language-learning infants lies in the resources they can use: in addition to the various types of cues to word boundaries, the most important resource available for adults is an existing lexicon that can be thought of as containing almost all words in the language – speech segmentation thus becomes an issue of decomposing a received utterance into a number of existing words in the lexicon that best match the input; whereas infant language learners start with an empty lexicon and also have to segment the speech input into words or word-like units. The infants seem to face a much more difficult task: they have to attempt a similar segmentation task with no comparable resources at all at the early stage of lexical learning – they may not even know there exist words in their mother tongue – and also they have to infer some cues to “word” boundaries for later use. What leads them to become aware of the existence of words in general and to extract individual words from fluent speech in particular?

A possible answer is that there may not be any particular thing(s) in speech signals that lead to this kind of awareness, but only that the human infants have an innate mechanism to derive a least-effort representation for the input data they encounter, and words happen to be the pieces of building blocks in such a representation. When the infants have more and more such pieces to a certain level, we call these entities *words*. The evolution of a lexicon can be rather dynamic, in that some old pieces (e.g., the multiple word collocations that were once recognised as individual words) may be dropped and some new pieces must be added, in order to achieve a least-effort representation for more and more new data. This dynamic lexicon finally comes to stabilise, relatively though, at a state in which it would not need any radical change to reach the least-effort representation while more and more upcoming input has either already been seen before or can be decomposed into fragments of the seen data. This stabilisation process seems to be a plausible process for an infant's lexicon to converge to the adult lexicon.

The hypothesis that human infants have an innate mechanism for language learning to derive the least-effort representation for language data follows the observation that language phenomena comply with the least-effort principle. In fact, it is rather straightforward to define the least-effort (or, the most economic) representation for a given set of data, following the *minimal description length* (MDL) principle [156, 157] – a popular approach to utilising an approximation of Kolmogorov complexity [166, 99, 112] to do inductive inference – and also formulate computer learning algorithms to realise this lexical learning strategy for the purpose of examining what performance this strategy can achieve on naturally-occurring language data. From this perspective, word discovery is viewed as a process of receiving input data and storing them in a representation as compact as possible, and the results of such leaning are the structures in the resulted representation – we call them words.

However, the research in this direction is beyond the scope of this paper. In this section, we will focus on the cognitive aspects of young infants' lexical learning. In particular, we review psycholinguistic studies on how infants exploit various cues to facilitate their inference and determination of possible word boundaries. As summarised in a recent review paper [86] by Jusczyk, the major cues used by infants in speech segmentation for extracting words from fluent speech to develop their lexicons include *prosodic* cues, *allophonic* cues, *phonotactic* cues and *statistical* cues (or distributional regularities). A large volume of background information on pre-linguistic infants' perceptual sensitivity to these cues can be found in Jusczyk's remarkable monograph [84] with thorough discussions. In the sections below, we will discuss the utility of these cues in young infants' lexical acquisition.

5.2.1 Prosodic Cues

Human infants are known to have special capacities for speech perception from a very young age – some are innately endowed and some are acquired after birth. In particular, they are remarkably sensitive to prosodic information, for example, the rhythm, in the fluent speech of their mother tongue. No doubt this perceptual sensitivity has to do with the prenatal exposure to speech sounds in the intra-uterine environment – this environment functions to protect the foetus from exogenous sounds by blocking sounds of high frequency (above 250 Hz) but letting the low frequency sounds be transmitted to the foetus' inner ear with little reduction of sound pressure [1]. Based on their prenatal experience of hearing, new-born babies are able to discriminate their mother tongue from other languages, based on their detection of the rhythmic distinction across languages [122, 134].

There appear to be two trends in infants' development of their awareness of linguistic units of various sizes, ultimately towards words. One trend develops from smaller to larger units, that is, the infants first learn the smallest discriminative speech units, i.e., *phones*, and their categorisation, resulting in the infants' awareness of *phonemes*, in their native language. Afterward, they pick up the structure of *syllables*, usually consisting of several phonemes in a hierarchical structure, and then, how syllables combine with each other to form words. Although these intra-word speech units carry certain suprasegmental information (e.g., tones over syllables), in general they do not provide, individually, much useful information about word boundaries. Allophonic cues are indeed an exception (see the next section for discussion).

Another trend is that the infants detect the existence of *clauses* (or *utterances*) as linguistic units in speech, and then *phrase*, and then words. The temporal progression of this trend is rather clear. It was first found in [75], using the head turn preference procedure with a pause insertion technique, that 7- to 10-month-olds demonstrated a preference for whole clauses over interrupted ones (i.e., ones with a pause inserted in the middle), although they were too young to understand the meaning of each clause. This result was interpreted as due to the infants' sensitivity to the prosodic demarcation of linguistic units such as clauses. Later, infants as young as only 4.5 months old were reported to have a similar preference for clausal prosody [83]. More strikingly, evidence is further given in [114] that 2-month-olds have a certain capacity to use clause prosody to organise and remember phonetic properties of words, e.g., *rat* versus *cat*. The infants appeared to remember speech information in a word better in natural clause prosody than in an isolated word list, as indicated by their responses measured in terms of their high-amplitude sucking (HAS) rates. It is also shown that during the age between 6 and 9 months, infants are developing their sensitivity to sub-clausal units such as prosodic phrases, as reflected in the fact that the 9-month-olds, but not the 6-month-olds, demonstrated a preference for listening to passages with pauses inserted at phrase boundaries rather than to passages with pauses inserted within phrases [91, 64, 84]. By the age of 11 months, but not before 9 months, infants appear to have developed their sensitivity to word boundaries in fluent speech to such a degree that is testable by the pause insertion technique [132].

Also, it is reported in [164], interestingly, that even new-born babies of 1- to 3-days old were observed to be able to discriminate between lexical and grammatical words in spontaneous infant-directed speech, relying on their perceptual sensitivity to constellations of acoustic cues, including some salient prosodic characteristics of the words, for example, grammatical words usually have a shorter vowel duration, weaker amplitude, simpler syllabic structure, and so forth.

From the above evidence, as a whole, we can see that the preverbal infants' perceptual ability to receive prosodic information may be in place from a very young age, if not from birth, and can be utilised to do speech segmentation and word discovery (although whether they have a sharp enough sensitivity to various kinds of prosodic cues so as to make use of them at the age around 7.5 months old when they start to recognise words in fluent speech [87, 86] still remains a question for the moment – see the discussion below). But what kinds of prosodic information (or cues) do the infants use to discover words in fluent speech? How effective are such cues and what problems may be caused by (over)using these cues? When problems are caused by one strategy of exploiting cues, how do the infants come up with a new strategy to utilise more information to overcome the deficiency of the old strategy?

According to Jusczyk [86], English speaking infants have a speech segmentation competence like a native adult around their second birthday, in terms of speed and accuracy. Many speech segmentation strategies are developed during the second half of the first year, and the skills are further improved in the second year. Utilising prosodic cues is just one of the strategies in the initial phase of lexical learning.

The term *prosody* denotes, in general, the suprasegmental attributes of speech sound, including pitch, stress, accent, duration, tone, intonation, rhythm, pause, etc., usually resulting from certain patterns of change of sound intensity and frequency. The recent studies on how prosodic cues facilitate pre-linguistic infants' lexical learning focus on the rhythmic (or metrical) structures of a few languages, e.g., English. Linguistically, *rhythm* refers to the harmonic succession of sounds, in particular, certain regular periodicity of some sound attribute(s), that reflects the musical flow of speech in a language. In English, for example, the rhythm is the patterns of alternation between strong and weak syllables, where a strong syllable (whose vowel is non-reduced) is known as a *stress*. A *stress foot* in English consists of a stressed syllable and a following unstressed syllable (which contains a reduced vowel), if present. The typical English stress feet generally have a *trochaic* stress pattern, e.g., *table* and *parent*, as opposed to an *iambic* stress pattern, each being a weak-strong syllable pair, e.g., *guitar* and *device*.

In order to be able to make use of prosodic cues in speech segmentation, infants must first have certain sensitivity to the cues. Inspired by the findings in [40] that most strong syllables are at the word onset in spoken English and hence the predominant stress patterns in English are trochaic, Jusczyk and his colleagues investigated whether American infants are sensitive to this typical prosodic property of English words [88]. They presented a list of trochaic or iambic bisyllabic words to 6- and 9-month-old infant subjects in experiments, and found that the 9-month-olds listened significantly longer to the trochaic than to the iambic words, whereas the 6-month-olds did not show any preference. In addition to [88], other subsequent studies also give indications that it is in the period of 6 to 9 months of age that English-speaking infants have developed their sensitivity to the predominant word stress patterns in their mother tongue [129, 128].

The infants' sensitivity to this rhythmic characteristic of English words appears to provide them with a basis for applying the metrical segmentation strategy (MSS) to start speech segmentation for word discovery. A number of interesting investigations were conducted by Jusczyk and his co-workers [77, 139, 84, 85, 93] on how English-learning infants of 7.5 months make use of the predominant stress pattern in the language to segment words from fluent speech, following the MSS. They first tested whether the infants could detect words in the predominant strong-weak stress pattern in fluent English speech. In the experiment, the infants were familiarised with a pair of target words (e.g., "hamlet" and "kingdom", or "doctor" and "candle"), and then presented with four test passages of a few sentences – two of these passages each carried a target word in each sentence and the other two did not. The experimental results showed that the 7-month-old infants listened longer to the passages containing the target words than the other passages, suggesting that they detected the familiarised strong-weak bisyllabic words in the passages.

However, there is another possibility in this experiment, that is, the infants might simply match the initial strong syllables of the target words, e.g., "doc" in "doctor" and "can" in "candle", to those in the passages. To eliminate this possibility, the experiment was changed to familiarise the infants with the first syllables of the target words and then present them with the test passages. This time, however, the infants' listening time showed no preference for the passages with the target words, suggesting that in the previous experiment the infants did detect the trochaic target words as a whole in the passages by whole-word matching, instead of matching the initial strong syllables of the words.

We know that the infants of this age follow the MSS to recognise trochaic words. How did they deal with words of the opposite stress pattern, namely, the weak-strong bisyllabic words? Jusczyk and his co-workers repeated the first experiment with the following change: they familiarised the infants with a couple of iambic target words such as "beret" and "device" or "surprise" and "guitar", and then presented to them four passages, two of which each contained a target word in each sentence and the other two, known as control passages, did not. This time the infants did not show any listening preference, suggesting that they did not detect any target words in the test passages.

What happened? The researchers guessed that since the infants were supposed to follow the MSS, it was possible that they inserted a word boundary at the middle of the weak-strong words that they heard in the familiarisation. To test his possibility, two more experiments were further conducted with iambic target words: the test passages were designed in a way such that each target and control word was followed by a particular unstressed word, e.g., "guitar is" and 'device to", to see how the infants would react to them. In one experiment, the subjects were familiarised with the iambic target words and then listened to the test passages – the result: no listening preference was detected. In the other experiment, the infants were familiarised with bisyllabic non-words of trochaic stress pattern such as "tar is" and "vice to", and then presented with the same test passages – this time, the subjects did listen significantly longer to the test passages with the target trochaic non-words!

These experiments all together demonstrate that English-learning infants do follow the MSS to identify strong syllables as word onsets at the time when they start to do speech segmentation at around 7.5 months old.

An experiment reported in [53] also gives further evidence for 9 month old infants' possible use of the trochaic stress pattern to segment English speech input into word-level chunks, because the subjects in the experiment, after hearing various speech inputs in the familiarisation phase, could distinguish the trochaic syllable pairs embedded in the four-syllable input from novel trochaic pairs in the test phase, but did not make this distinction for the iambic targets and novel iambic distracters – this result suggests that the infant subjects could recognise the previously heard trochaic bisyllabic sequences in the speech input as familiar lexical units that stand out. Notice, however, that the four-syllable speech input is not naturally-occurring speech data from fluent speech.

However, entirely relying on the metrical segmentation strategy to do speech segmentation does lead to mistaken results – the infants will always miss the iambic bisyllabic words. It has not been clear what kind of strategy the infants would use to remedy the problems caused by the MSS strategy, although it is known that the infants at 10.5 months of age have gained the ability to recognise iambic words in fluent speech [86]. It is reasonable to guess that infants of this age may resort to a constellation of available cues. Also, what strategy an infant would exploit to integrate multiple cues is also unclear, although there is evidence that, for 9-month-old infants, when both prosodic and phonotactic cues are available but conflict with each other, prosody overrides phonotactics [118].

Specifically for rhythmic cues, we have a few questions to ask: Where are they from? How do the English-learning infants get sensitivity to them and acquire the sense (or knowledge) that the stressed syllables are more likely to align with word onsets? Do they have such knowledge before knowing any words, or they learn such knowledge through the experience of knowing words?

There is an observation [84] (pp.108) that English-learning infants may learn the rhythmic cues to word onsets and the MSS strategy from their experience of listening to isolated words, in particular, English first names (e.g., Peter, Tommy, David, etc.) and diminutive forms of words (e.g., doggie, cookie, kitty, daddy, mommy, etc.) that adults repeatedly use in isolation around the infants, mostly for the purpose of catching the infants' attention. It is reported that the strong-weak stress pattern is common in English first names [41] and that infants start to recognise their names when they are around 4.5 months old [115]. From 4.5 to 7.5 months old, infants have quite a lot of time to experience the isolated trochaic words such as the diminutives and their names and those of their close relatives and caregiver(s), and consequently, develop their sensitivity to the rhythmic patterns of English words. This observation lends support to our argument that cues for words should be learned from known words.

5.2.2 Allophonic Cues

Allophonic cues for speech segmentation are the phonetic variants of some phonemes in a language that correlate with word boundaries. Each acoustic realisation of a phonetic variant of a phoneme is known as an *allophone*. For example, as noted in [31], the /t/ at the onset of a word in English, e.g., as in "tap" ([t^{h}]), has a different pronunciation from the /t/ in other places within a word, e.g., as in "stop" and "hat". The possibility that allophonic variants of phonemes can signify word boundaries was noted in a number of early studies in 1960-70's, such as [110, 170]. A frequently quoted example in previous discussions is the pair "nitrate" *versus* "night rate": in the former the first /t/ is aspirated, released and retroflexed and /r/ is devoiced, suggesting that /tr/ forms a cluster that only appears in a within-word context; whereas in the latter the first /t/ is unaspirated and unreleased, and the /r/ is voiced – this /t/ and /r/ together indicate a word boundary in between. One more example to illustrate that allophones really can signify word boundaries is the pair "nice top" *versus* "nice stop", as given in [84].

We know that infants perceive speech sounds categorially from a very young age – in general, they hear sounds in terms of their phonological functionality: speech signals with acoustic distinction but no distinctive phonological significance are heard as the same sound, namely, a *phoneme*. However, this categorial perception does not by any means deprive the infants of the ability to detect the acoustic difference between individual instantiations of the same phoneme in conversational speech. It is demonstrated in [78] that 2-month-old infants could detect the difference between the allophonic variants of /t/ and /r/ in the distinctive pairs “nitrate” and “night rate”.

However, this does not necessarily mean that 2-month-old infants have the ability to make use of allophonic cues to segment speech into words. Whether the allophonic cues can be used by language-learning infants to facilitate speech segmentation depends on the following conditions, discussed repeatedly in Jusczyk and his colleagues’ recent study [92]:

1. Allophones are an orderly manifestation, instead of random acoustic variants, of phonemic contrasts;
2. Allophones have a distributional correlation with word boundaries;
3. Infants are able to discriminate the allophones from each other;
4. Infants are sensitive to the distribution of allophones within words;

The first three of these conditions are known to have had support from previous relevant phonetic and phonological studies, as discussed above. The psycholinguists’ task is to acquire empirical support for the infants’ sensitivity to allophonic cues and their actual use of the cues in speech segmentation.

There are four experiments reported in [92] that are aimed at testing whether and by which age English-learning infants can have sensitivity to the distribution of allophonic cues within words. In each experiment, 24 infants of 9 months old were tested with 4 words, two target words and two control words. Target words carried allophonic cues of interest, such as “nitrates” and “night rates” – the allophonic cue was the only difference between them; whereas control words carry no cues, but differ from each other significantly in other ways, such as “hamlet” and “doctor”.

In the first experiment, the infant subjects were each familiarised with two words: a target word and a control word. Each of these words was presented in isolation repeatedly to each subject until some familiarisation criterion (e.g., listening time accumulated up to 30s) was met. Then, each of the subjects was tested for each word with a passage of six sentences, each of which contained the particular word once – i.e., the word was repeated six times in the passage. In the test for each subject, all four passages were heard: two for familiar words – one carried an allophonic cue and the other did not – and the other two for unfamiliar words – also one carried an allophonic cue and the other did not – as below, for example:

	With	Without
	allophonic cue	allophonic cue
Familiar	*night rates*	*doctor*
Unfamiliar	*nitrates*	*hamlet*

An infant in the experiment was either familiarised with “night rates” and “doctor” or with “nitrates” and “ hamlet”, and then listened to all four passages. The experimental results based on the analysis of the mean listening time to the four passages by each infant indicate that the difference of listening time for the familiar and unfamiliar items is significant for “doctor” and “hamlet” – the pair without allophonic cue, but is not significant for “night rates” and “nitrates” – the pair with allophonic cue. This result suggests that the 9-month-olds did not use the allophonic cues to match the target words they heard during the familiarisation to the test passages containing the corresponding target words.

The second experiment used two monosyllabic words, namely, “night” and “dock”, instead of bisyllabic words, for the familiarisation task, and then the subjects were tested with “nitrates’, “night rates”, “dock” and “doctor” passages, to see if the memory demands for the bisyllabic words in the previous experiment was what blocked the 9-month-olds from using allophonic cues. However, the result turned out to show that the allophonic cue in “night” was not used to match the word either to the “night rates” or the “nitrates” passage.

The next experiment tested whether the infants could recognise the word “night” in the “night rates” passage, to clear the doubt caused by the previous experiment. The experiment repeated most of the previous one, except that each occurrence of “night rates” in the test passage was changed to either “night time”, “night games” or some other “night X” item, for the purpose of introducing distributional regularities. The experimental outcomes showed that the infants listened longer to the “night X” passage than the “nitrates” one, suggesting that the infants did recognise “night” in the passage. Based on this result, we can infer that the infants took “night rates” as an individual lexical item different from, and thus independent of, “night”.

These three experiments so far gave a clear indication that the 9-month-old English-learning infants were not sensitive to allophonic cues and therefore could not make use of them to extract familiar words from fluent speech. However, would older infants possibly build up their sensitivity to allophonic cues later, say, at the age of 10.5 months old? It is known that infants of this age have developed many abilities to detect words in fluent speech, for example, they can detect words in the iambic stress pattern (beyond their detection of trochaic words since 7.5 months old with the aid of MSS) [77], and can also detect the interruption of words, in either the trochaic or iambic stress pattern, by a pause [132], suggesting that they have had a very good awareness about the well-formedness of individual words by this age.

When the fourth experiment was conducted by repeating the first experiment with 10.5-month-old infants, the result was that the difference of listening time for the familiar and unfamiliar items is significant not only for the pair without an allophonic cue, namely, “doctor” and “hamlet”, but also, more interestingly, for the pair with an allophonic cue, namely, “night rates” and “nitrates”. Comparing this result with that of the first experiment with 9-month-old infants, we can conclude that the English-learning infants have developed a testable sensitivity to allophonic cues between 9 and 10.5 months of age.

Notice, however, there is no direct evidence in [92] showing that the infants actually use the allophonic cues to do speech segmentation, although the experimental outcomes are no doubt consistent with, and even supportive of, this possible use of allophonic cues. More research is needed in this direction to obtain proof as well as other evidence about the effects of using allophonic cues in speech segmentation for lexical discovery.

5.2.3 Phonotactics

Phonotactics is concerned with what sound sequences (in particular, consonant sequences) are permissible and what are disallowed in a word or in some particular positions (e.g., the onset and offset) of a word in a language. For example, sound sequences such [db], [kt] and [zb] are common as the syllable onset in Polish words, but never in English or Chinese words. Also, the [mr] sequence is also disallowed as a syllable onset in both Chinese and English words; whereas [st] is a frequent sound sequence in English (e.g., "stop", "stone", "best", "least"), but not in Chinese.

Native speakers' phonotactic judgments appear to be relative to the co-occurring frequency of sounds in their language experience that have shaped their intuition about what sound sequences are allowable and what are unlikely in their language. The extreme end of being unlikely is "disallowable". Disallowable sound sequences are related to low co-occurring frequency, in particular, the frequency zero that indicates a sound sequence is never observed in the language. Usually we use the term *phonotactic constraints* (or *patterns*) to refer to the disallowed sound sequences in a language. Because of the close relation of the legality and illegality of this kind to the probability of co-occurrence, or more precisely, to the distributional regularities, they are also known as *probabilistic phonotactics*, e.g., as in [171, 117].

It is easy to infer that once a phonotactic pattern appears, it signifies a word boundary. For example, once the phonotactic sequence [$\cdots$mr$\cdots$] is heard in English or [$\cdots$st$\cdots$] in Chinese, it is quite clear to native speakers that a word boundary must appear in between. Fluent speakers, including adults, adolescents and even three- to four-year-old children, are highly sensitive to phonotactic regularities [155, 172, 171], in that they can respond faster to phoneme sequences (either words or non-words) of high-frequency than ones of low-frequency. More interestingly, many high-frequency non-word phonemic sequences are judged by children as more likely to be words than some real words consisting of some rare but legal phonemic sequences, and the children also pronounce such "words" more accurately than the "non-words" [123]. There is evidence that adult speakers exploit their sensitivity to phonotactics to facilitate word segmentation from fluent speech [120].

From the computational perspective, in addition to an early proposal by Church [31] to make use of phonotactic information to facilitate computational lexical processing, it has been demonstrated in Brent and Cartwright's studies [25, 18] on computer simulation of lexical learning that phonotactic information can be used to significantly enhance the performance of unsupervised lexical learning.

Language-learning infants also appear to have a proper sensitivity to phonotactic well-formedness of words at a very young age. It is reported in [90] that between a list of words with phoneme sequences legal in English but not in Dutch and a list of words with phoneme sequences legal in Dutch but not in

English, English learning infants of 9 months old listened longer to the former, whereas Dutch infants of the same age preferred the latter, and both English and Dutch younger infants, of 6 months old, did not demonstrate any significant preference. Also, it is observed in another study [94] that English infants of 9 months old, but not of 6 months old, demonstrated a greater preference for monosyllabic non-words with phonotactic sequences of high-frequency (e.g., *chun*) than the ones with phonotactic sequences of low-frequency (e.g., *yush*).

Furthermore, the young infants also appear to have a certain awareness of the legal position for some phonotactic sequences within words. For example, in Dutch, there are some typical consonant sequences for word onsets (e.g., [br]) and some for word offsets (e.g., [rt]). It is reported in [61] that Dutch infants of 9 months old, but not of 4.5 and 6 months old, can discriminate monosyllables with consonant sequences in permissible positions (e.g., [bref] and [murt]) from those with consonant sequences in impermissible positions (e.g., [febr] and [rtum]), according to their listening preference measured by listening time.

The young infants' sensitivity to phonotactics, reviewed above, seems to provide a nice basis to support the hypothesis that they can exploit phonotactic information to detect word boundaries [23]. The phonotactic sequences, in particular the cluster of between-word sequences, are known to carry probabilistic information about word boundaries.

Interestingly, a couple of recent studies [118, 117] give further evidence that young language learning infants do make use of phonotactic cues to segment fluent speech into words based on their sensitivity to the alignment of phonotactic sequences to word boundaries. The first experiment in [118] gave confirmation that 9-month-old infants prefer, in terms of their listening time, a two-consonant sequence C·C of the within-word cluster over that of the between-word cluster to appear in a bi-syllabic non-word CVC·CVC (e.g., "nongkuth" versus "nongtuth", where the two clusters are intentionally selected so as to have equally high probability in terms of their frequency in the Bernstein child-directed corpus [6] so that the two clusters only differ in their likelihood of being within a word or at a word boundary). The major findings of the other experiments in [118] are more worth noting: the infants also listened longer to the bi-syllabic non-words with between-word phonotactic sequences showing up in between the two syllables than those with within-word sequences showing up in between, under either of the two following conditions: (1) the second syllable is stressed; or (2) the first syllable is stressed plus a 500-ms pause of silence is inserted in between the two syllables. Recall that language-learning infants of this age tend to assume a stressed syllable is the onset of a word, and also notice that the 500-ms pause was inserted as a delimiter between words. Thus, we can see that the experimental results gave not only an indication that the infants felt that the between-word phonotactic sequences were more natural aligned with word boundaries, but also gave supporting evidence for the view that the infants use the phonotactic regularities to locate potential word boundaries.

More remarkably, stronger evidence was given in [117] that the infants of 9 months old did exploit probabilistic phonotactics to segment words from fluent speech. Three experiments were conducted in [117]: the subjects were first familiarised with a passage of six sentences with two target words – one of which was a real word (e.g., "gaffe") and the other a non-word (e.g., "tove") – either one of the targets was surrounded by a phonotactic sequence of low within-word (i.e., and high between-word) probability[2] at both the onset and offset or at only the onset or offset of the word, respectively, in the three experiments, and then tested, in each experiment, with the two targets and two control stimuli – one of which was a real word (e.g., "pod") and the other was a non-word (e.g., "fooz"). During the familiarisation phase, the chance for the real word and the non-word to be surrounded by phonotactic cues was half by half in each experiment. Accordingly, there were two types of targets in the familiarisation phase: "P-cues present" (the one with some phonotactic cues) versus "P-cues absent"(the one with no cues). The experimental results on 24 infants were as follows: they listened significantly longer to the P-cues present target than the others, and their listening time for the P-cues absent target was not significantly different from the two unseen control stimuli. This means that the infants could recognise the target words they heard in the passage that were surrounded by phonotactic cues but could not detect the target word that they heard with no phonotactic cues. The three experiments give a clear indication that the infants did segment words from fluent speech using the probabilistic phonotactics, no matter whether the phonotactic cues appeared at one side or both sides of the target word's edges.

5.2.4 Distributional Cues

Human infants are also found to be sensitive to the *statistical regularities* embedded in the speech input, in particular, the co-occurring patterns. Since such statistical regularities give hints to word boundaries, we also refer to them, alternatively, as *distributional cues.* It is argued in [139] that the reason why English-learners do not misidentify "can" as a single word in a passage talking about "candle" has to do with their sensitivity to the distributional cues.

How *transitional probability* may have an effect on infants' segmentation tendencies is studied in [69]. In the study, 7-month-old infants were first trained with a conditional headturn procedure to discriminate two separate syllables [ti] and [de]. Then they were tested on how well they can identify these two target syllables in various two-syllable contexts. A target syllable may combine with a two-syllable context in one of the following three ways:

1. The contexts are *invariant order* strings, e.g., [koga], but the target can appear at either edge, e.g., [tikoga] and kogati];
2. The contexts are *variable order* strings, e.g., [kogati] and [gakoti];
3. The contexts are two identical non-target syllables, referred to as *redundant* strings, e.g., [kokoti];

[2] The probability is also estimated in terms of the frequency in the Bernstein corpus [6].

Experimental results showed that the infants' performance on identifying the two target syllables in the invariant order contexts was the best. The researchers concluded that since the two context syllables always co-occurred in the same order with a very high transitional probability, this coherence or distributional property enabled the infants to group the two context syllables into one unit. That is, the two fixed-order context syllables and the target syllable were, respectively, segmented into individual units. Similarly, the redundant context should lead to similar results, but did not. The researchers explained that the two identical syllables lacking coarticulatory cues might be the factor that led the infants not to group them as one unit.

This line of investigation was subsequently extended in [127] to examine how the distributional cues and rhythmic cues would interact to facilitate speech segmentation. In the experiments, the infants were first trained to discriminate the target syllables [ti] and [de], as in [69], and then tested to see how capable they were of identifying the syllable in various two-syllable contexts:

1. The context syllables were always trochaic and fixed-ordered – both rhythmic and distributional cues available;
2. The context syllables were always trochaic but their order varied – rhythmic cues available but distributional cues not;
3. The context syllables were not ordered in any way – neither rhythmic nor distributional cues available.

The experimental results showed that the infants had a better performance in identifying the target syllables under the first two conditions than the last one, and more interestingly, when the testing lasted for some longer time, the infants' performance under the first condition was significantly improved but such improvement did not take place under the second condition. These results suggest that the infants were able to integrate the distributional and rhythmic information to do speech segmentation when they were available in the input.

Later, this line of investigation was further extended in [129] with various experimental techniques, yielding the finding that 9-month-old infants' performance was significantly better when both sequential (i.e., distributional) and rhythmic information were available in the context syllables than when only one of the two was available. In contrast, 6-moth-old infants performed equally well when both sequential and rhythmic or only rhythmic information was available. These results suggested that infants, at the age of 9 months old, could integrate different sources of information such as the distributional and rhythmic cues, when available, to do segmentation; whereas at 6 months old they could only use the rhythmic information.

More direct proof that young infants use sequential statistics in speech segmentation and word learning is presented in [159]. In the experiments, 24 8-month-old infants were first familiarised with a continuous speech stream of only four tri-syllabic artificial nonsense "words" (e.g., *tupiro*, *golabu*, *bidaku* and *podati*) in a random order, generated by a speech synthesiser as a consonant-vowel sequence at the rate of 270 syllables (i.e., 90 words) per minute with no pause, no stress differences or any other acoustic or prosodic cues to the word boundaries. In the speech stream, the only cues to word boundaries were the transitional probabilities, which were higher within words (1.0 in all cases) than across words (0.33 in all cases). In the first experiment, after listening for 2 minutes, i.e., 180 words in total, the infants were tested with repetitions of two tri-syllabic words (e.g., *tupiro*, *golabu*) and two tri-syllabic non-words (e.g., *dapiku* and *tilado*). Each syllable of the non-words appeared in the stream, but no bisyllabic sequence in the non-words was ever heard by the infants. The test results showed that the infants made a significant discrimination between the words and non-words by listening longer to the non-words – the result of novelty preference, suggesting that the infants could recognise the words they heard during the familiarisation, with the aid of transitional probability. In the second experiment, the familiarisation was similar but the test was different: the infants were tested with two tri-syllabic words and two tri-syllabic part-words. The part-words were tri-syllabic sequences each crossing a word boundary in the speech input and all having been heard by the infants during the familiarisation. Again, the infants listened longer to the part-words, suggesting that the infants judged them as novel items, in contrast to the words, which were familiar items – the infants must have learned the words with sufficient specificity and completeness, otherwise they would not have treated the part-words crossing a word boundary as unfamiliar. The infants' performance in this more difficult discrimination task suggested that they were able to extract sequential statistic information from the input and use it to segment the word-like units out of the speech stream – this can be attributed to nothing else than transitional probability that the infants somehow computed based on their listening experience. This evidence indicates that language-learning infants have a learning mechanism to learn statistical information from speech input to facilitate lexical learning at a very young age.

However, the above experiments do not tell us reliably what particular kind of statistical computation the infants really performed for the purpose of segmenting words from the speech input, since in the experiment both transitional probability and the co-occurring frequency of words were higher than those of the part-words. To tackle this specific issue, the same group of researchers adjusted the design of the previous experiments in several ways, as reported in [2]. First, the speech stream of four tri-syllabic words in a random order was generated in a way such that two words were twice as frequent as the others, instead of all words being equi-probably generated. More specifically, in the familiarisation phase when the infants heard the speech input for 3 minutes (i.e., 270 words in total), two words appeared 90 times each and the other two 45 times each. Second, the words and part-words used in the test were selected so that they all had the same frequency (i.e., 45 times) in the speech stream. The only difference between the test words and part-words this time was the transitional probability: The transitional probability of any bi-syllabic sequence within a test word remained 1.0 and that within a test part-word was either 0.5 (if the two syllables in question crossed two words) or 1.0 (if the two syllables were within a word). The experimental result was that the infants' novelty preference for part-words was significant, indicating that it is the transitional probability, not the co-occurring frequency, that the infants computed during the statistical learning of words.

Unfortunately, viewed from a computational perspective, the belief that a learner would follow a piece of distributional information such as transitional probability to determine the word boundaries in human speech is oversimplified and even naive. The artificial speech stream used in the experiments reported in [159, 2] is an extreme case: all words were of the same length, no word embedding, no words shared any common syllable, all transitional probabilities within words were 1.0 and those between words were 0.33 (or 0.5), etc. What is actually behind the naive idea that the learners compute transitional probabilities to determine word boundaries is the idea that the learners would naively infer that the lower points (or local minima) of transitional probability are word boundaries – computational studies have shown that only about half of all words in a language like English can be discovered this way [16]. The main evidence against this naive idea is that there are so many local minima of transitional probability appearing within a word that would give false alarms for word boundaries, and also, there are many word boundaries at which the transitional probabilities are not local minima. More interestingly, there are so many kinds of statistical computation that may give an indication to word boundaries with a certain reliability, how can we be sure it is exactly the transitional (i.e., conditional) probability that plays the role of guiding the infants to word boundaries? More specifically, as far as transitional probability is concerned, how do we know whether the learners use the transitional probability of a syllable given the preceding one syllable or two or three or more, or use the transitional probability of two or three syllables given the preceding one syllable or two or three or more? There are so many possibilities out there, and it appears impossible to follow the approach of [2] to determine which one is used by lexical-learning infants by eliminating all other possibilities.

At this point, we have a number of questions to ask: What is the rationale behind the idea that the learners have to use transitional probability? Is there any theoretical significance or preference for using transitional probability over using other statistical measures? From what theoretical assumption(s) can we arrive at the conclusion that the learners are likely to, or have to, use transitional probability? So far, we have no satisfactory answers to any of these questions from the psycholinguistic studies of lexical-learning infants' statistical learning ability. What we see is that transitional probability can be relevant to some extent, just as many other types of statistical measure may be relevant. In short, what we know for sure is that the infants do compute something statistically based on their speech experience for the purpose of lexical learning, but we are not sure what they actually compute – it is yet to be explored. More importantly, the cognitive principle(s) underlying this computation is (are) also to be investigated, especially, in a more principled way.

To explore the underlying principles beneath a lexical learner's statistical computation is one of the major motivations for proposing the least-effort approach to the computational studies of lexical learning in Section 5.2. Within this approach, the theoretical assumption is clearer, more straightforward and better rooted: a language learner learns linguistic regularities, including words, as speech patterns from naturally-occurring language data that are known to be generated by human speakers following the least-effort principle [179, 180]. It is reasonable, and necessary, to assume that a learner must have a least-effort learning mechanism to accommodate the data, because when the learner has learned a language, he or she also produces new utterances in the same way following the least-effort principle. Computationally, this least-effort learning mechanism is formulated as a strategy to seek for the least-effort (i.e., minimal-cost) representation for the observed language input in terms of storage space. Lexical learning can be viewed as an application, or an implementation, of this learning mechanism that is specialised for inferring lexical items from language data: it searches for a lexicon (i.e., a set of lexical items such as words) consistent with the data that can represent the data with a minimal cost. Computationally, the cost is measured in the number of bits that are necessarily used to present the data and the lexicon itself. It is unrealistic to expect a computer to carry out all human learning tasks, but the implementation of this computational strategy as a program can no doubt help us explore how far this learning mechanism can go in the direction of learning word-like lexical items from real language data. Some representative studies in this direction can be found in [144, 47, 16, 17, 98], among many others.

6 Summary and Discussion

In this article we have reviewed the cognitive, in particular psycholinguistic, aspects of recent studies on lexical acquisition by very young human infants. From the cognitive aspect, we can see that preverbal infants are well prepared, perceptually, for lexical learning, in that they can detect individual phonemes (and their acoustic variants), syllables, prosodic marking of various structures (including clauses and phrases), rhythmic patterns, etc., in the speech stream of their native language from an early age. This provides them with a ground to start lexical learning – they can identify the syllables and phonemes, which are the basic building blocks of words in speech utterances. We can see a clear scenario for lexical learning as this: a language learning child receives a stream of these building blocks utterance by utterance and induce, with little supervision, the words from this stream with all aids useful and available, including prosodic (rhythmic) characteristics, allophonic cues, phonotactics, distributional regularities, and perhaps some other means we do not know yet.

A problem with the cue-based approach to lexical acquisition is that the question of how children learn lexical items starting from an empty lexicon is transformed into the question of how they learn (or develop) the cues. That is, if we were asked how infants learn words, we could have a nice answer that they learn words with the aid of some cues; however, if we were further asked how do they learn the cues, we are, unfortunately, yet to look for a reasonable answer. More interestingly, before knowing any words, how can the infants know anything that can be cues for words? They are not born with such knowledge – at the very beginning, the infants are not even aware of the existence of words in their language.

Without knowing the existence of words, why do they learn words? Why not some other recurring patterns or some arbitrary strings in the speech input? In other words, what kind of force drives them to learn words, instead of any things else? A possible answer to this question is that, as stated before, the learning infants are not aware of what they are learning, nor the existence of words in their language; what they do in general is use an innate mechanism to accommodate the input data with an economic representation (or storage). Segmenting the input into chunks is a preferable approach for achieving this accommodation at an early stage of language development – we call this stage *lexical acquisition*. It is also reasonable to assume that human infants also have an innate statistical learning mechanism to detect the distributional characteristics in the input stream and guide the segmentation to arrive at a representation as economic as possible. When the infants perceive more input, they have more chunks and more experience in doing the segmentation. Gradually, the chunks converge to a stable set – during the convergence, some other properties (e.g., reference, meaning) may be attached to individual chunks, and the infants also develop some rudimentary ability to string up the chunks to produce their own utterances. Linguists call this stable set the *lexicon* and the individual chunks *words*.

Of course, this sketchy scenario does not mean to deny the existence of cues during lexical development. But it is important to recognise that the cues should be learned somehow from known words or from the learners' experience of lexical learning. The learned words can, in turn, be the most reliable "cues" for spotting the boundaries of new words. This looks like a more effective way to keep the snowball rolling, in comparison against a fixed set of cues. When the infants have developed their lexicon to a certain number of words, they may rely more on the existing lexicon for word segmentation and new word discovery than on the cues not so reliable.

However, the most interesting question we have asked about lexical learning is how the learners start to learn words without any cues available? Is it possible for a learner to learn words without any given cues and constraints but entirely relying on the speech stream as a sequence of sounds with intrinsic distributional regularities? This can be the real situation that a preverbal language learner has to face and overcome, in that they have to derive the useful cues by themselves. Before the derivation of any cue, however, they have nothing more than the distributional information from the input stream. This idea of basing the lexical learning solely on distributional information is at the core of the theory known as the *autonomous bootstrapping hypothesis* [15], in contrast to other bootstrapping hypotheses such as prosodic bootstrapping and semantic bootstrapping which assume that information about linguistic structures of one type (e.g., prosody, rhythm, etc.) may give some clues to initiate the acquisition of linguistic structures of another type (e.g., syntactic structures like phrases and clauses, and lexical items).

We have argued that all reliable cues and constraints on lexical boundaries must be derived (or acquired) through perceptual experience during the lexical learning process, instead of given *a priori* from other recourses before the learning process. It is logically reasonable that recognising something as a cue or constraint for words must be based on the learner's knowledge about words. That is, one must know some words first and then can induce reliable cues and constraints for words. The cues so learned can be applied to identify known words and discover new words in upcoming speech utterance. When more words are learned, the leaner may induce more cues and other types of knowledge to facilitate word identification and discovery, and also remedy any existing deficiencies in the old cues and strategies of learning. Gradually, they rely on known words more than any other cues.

Our argument is that it is illogical, if not inconceivable, that a learner could acquire any cues for words without any experience of knowing some words. And a certain number of words that a learner learns at the beginning of lexical acquisition are unlikely to be learned solely with the aid of some prosodic and/or allophonic cues. Instead, distributional information (in particular, the repetition and co-occurring patterns) may play a more essential role in helping the infants to start lexical acquisition. This is not to deny the usefulness of various kinds of cues in some later phases of lexical acquisition – their usefulness has been evidenced by so many psycholinguistic experiments – but to argue that distributional information is what the learning can rely on before the availability of other cues. Many cues, e.g., phonotactics, are actually strongly related to, and even derived from, the distributional regularities in the speech stream.

The autonomous bootstrapping hypothesis does not rule out the possibility of seeking help from other resources, in that any useful means available may be used. Basically, it assumes that learning should primarily rely on resources within the speech input data for lexical learning, i.e., the co-occurrence and distributional information. The critical point is that, even without external resources, lexical learning can succeed to a great extent in the discovery of new words if the learners have certain statistical learning capacity to detect distributional regularities in the input.

Recent achievements in computational studies on lexical learning have lent significant support to the autonomous bootstrapping hypothesis. Almost all computational models rely solely on distributional information, and many of them have shown an outstanding performance on lexical learning. For example, de Marcken's concatenative model [47] achieved a recall rate higher than 90%, and Brent's MBDP-1 algorithm [16], based on his probabilistically sound model, achieved an impressive performance with both precision and recall at the level of 70-80%. More importantly, Brent's recent progress in his research has demonstrated that a better model to utilise distributional information is more critical, and therefore closer to the underlying learning mechanism, than a poor model with the aid of heuristic cues such as phonotactics. It is also showed in [98] that an optimal segmentation of input utterances into chunks with the greatest *description length gain* (DLG), following the essence of the *minimum description length* (MDL) principle, recognises around 3/4 of the English words (tokens) in the original text transcript of the Bernstein corpus of child-directed speech [7]. All these successes in machine learning of a natural language lexicon indicate that the distributional regularities in the data may play a fundamental role in human lexical acquisition, and it is essential to assume that human infants have an innately endowed mechanism to utilise such statistical information in language learning. Although the operational details in machine learning are undoubtedly different from those in human learning, the fundamental mechanisms involved in human and machine learning to utilise the same information source to detect the same linguistic patterns or regularities should be highly similar in principle, and the studies on both sides should shed light on each other.

Acknowledgements

The author wishes to thank Yorick Wilks for his enthusiastic support, beneficial advice and various kinds of help that have enabled this study. Special thanks go to Randy LaPolla, Chunyan Ning, Haihua Pan and Jonathan Webster, for valuable comments and helpful discussions. Randy and Haihua's effort on proofreading is also gratefully acknowledged. Any remaining imperfections are the author's.

References

[1] R. M. Abrams, K. J. Gerhardt, and A. J. M. Peters. Transmission of sound and vibration to the fetus. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, and W. P. Smotherman, editors, *Fetal Development: A psychobiological perspective*, pages 315–330. Lawrence Erlbaum, Hillsdale, NJ, 1995.

[2] R. N. Aslin, J. R. Saffran, and E. L. Newport. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4):321–324, July 1998.

[3] E. Bates, I. Bretherton, and L. Snyder. *From First Words to Grammar: Individual differences and dissociable mechanisms.* Cambridge University Press, Cambridge, 1997.

[4] H. Benedict. Early lexical development: comprehension and production. *Journal of Child Language*, 6:183–200, 1979.

[5] J. Berko. The child's learning of english morphology. *Word*, 14:150–77, 1958.

[6] N. Bernstein. Acoustic study of mother's speech to language -learning children: An analysis of vowel articulatory characteristics. *Dissertation Abstract International*, 43-04B, 1982. [No. AAI8220909].

[7] N. Bernstein-Ratner. The phonology in parent child speech. In K. Nelson and A. van Kleeck, editors, *Children's Language: Vol. 6.* Erlbaum, Hillsdale, NJ, 1987.

[8] R. Bijeljac-Babic, J. Bertoncini, and J. Mehler. How do 4-day-old infants categorize multisyllabic utterances. *Developmental Psychology*, 29:711–721, 1993.

[9] L. Bloom. *Language Development: Form and function in emerging grammars*. MIT Press, Cambridge, MA, 1973.

[10] P. Bloom. Subjectless sentences in child language. *Linguistic Inquiry*, 21:491–504, 1990.

[11] P. Bloom. Syntactic distinction in child language. *Journal of Child Language*, 17:343–55, 1990.

[12] J. N. Bohanan, B. MacWhinney, and C. Snow. No negative evidence revisited: Beyond learnability or who has to prove what to whom. *Developmental Psychology*, 26:221–6, 1990.

[13] M. D. S. Braine. Children's first word combination. *Monographs of the Society for Research in Child Development*, 41, 1963.

[14] M. D. S. Braine. On learning the grammatical order of words. *Psychological Review*, 70:323–48, 1963.

[15] M. R. Brent. Advances in the computational study of language acquisition. *Cognition*, 61:1–38, 1996.

[16] M. R. Brent. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34:71–106, 1999.

[17] M. R. Brent. Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Science*, 3:294–301, 1999.

[18] M. R. Brent and T. A. Cartwright. Distributional regularity and phonological constraints are useful for segmentation. *Cognition*, 61:93–125, 1996.

[19] R. Brown. Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology*, 55:1–5, 1957.

[20] R. Brown. *A First Language: The early stages*. Harvard University Press, Cambridge, MA, 1973.

[21] R. Brown and U. Bellugi. Three processes in the child's acquisition of syntax. In E. H. Lenneburg, editor, *New Directions in the Study of Language*. MIT Press, Cambridge, MA, 1964.

[22] R. Brown and C. Fraser. The acquisition of syntax. In C. Cofer and B. Musgrave, editors, *Verbal Behavior and Learning: Problems and Process*, pages 158–201. McGraw-Hill, New York, 1963.

[23] P. Cairns, R. Shillcock, N. Chater, and J. Levy. Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33:111–153, 1997.

[24] S. Carey. The child as word learner. In M. Malle, J. Bresnan, and A. Miller, editors, *Linguistic Theory and Psychological Reality*, pages 264–93. MIT Press, Cambridge, MA, 1978.

[25] T. A. Cartwright and M. R. Brent. Segmenting speech without a lexicon: The roles of phonotactics and speech source. In *First Meeting of the ACL Special Interest Group in Computational Phonology*, pages 83–90. AssoSciation for Computational Linguistics, 1994.

[26] N. Chomsky. Review of B. F. Skinner's *Verbal Behaviour*. *Language*, 35:26–58, 1959.

[27] N. Chomsky. *Aspects of the Theory of Syntax.* MIT Press, Cambridge, Mass., 1965.

[28] N. Chomsky. *Language and Mind (enlarged edition).* Harcourt Brace Jovanovich, New York, 1972.

[29] N. Chomsky. *Lectures on Government and Binding.* Foris, Dordrecht, 1981.

[30] N. Chomsky. *Knowledge of Language: Its Nature, Origin, and Use.* Praeger, New York, 1986.

[31] K. W. Church. Phonological parsing and lexical retrieval. *Cognition*, 25:53–69, 1987.

[32] H. E. Clark and E. V. Clark. *Psychology and Language: An Introduction to Psycholinguistics.* Harcourt Brace Jovanovich, New York, 1977.

[33] J. Colombo and R. Bundy. A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, 4:219–33, 1981.

[34] R. P. Cooper and R. N. Alsin. Preference for infant-directed speech in the first month after birth. *Child Development*, 61:1584–1595, 1990.

[35] R. P. Cooper and R. N. Alsin. Developmental differences in infants attention to the spectral properties of infant-directed speech. *Child Development*, 65:1663–1677, 1994.

[36] S. Curtiss. *Genie: A psycholinguistic study of a modern-day "whild-child".* Academic Press, New York, 1977.

[37] A. Cutler. Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmana, editor, *Cognitive Models of Speech Processing: Psycholinguistic and computational perspectives*, pages 105–121. MIT Press, Cambridge, MA, 1990.

[38] A. Cutler. Segmentation problems, rhythmic solutions. *Lingua*, 92:81–104, 1994.

[39] A. Cutler and S. Butterfield. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31:218–236, 1992.

[40] A. Cutler and D. M. Carter. The predominance of strong initial syllables in the english vocabulary. *Computer Speech and Language*, 2:133–142, 1987.

[41] A. Cutler, J. McQueen, and K. Robinson. Elizabeth and john: Sound patterns of men's and women's names. *Journal of Linguistics*, 26:471–482, 1990.

[42] A. Cutler, J. Mehler, D. G. Norris, and J. Segui. The syllable's differing role in segmentation of french and english. *Journal of Memory and Language*, 25:385–400, 1986.

[43] A. Cutler, J. Mehler, D. G. Norris, and J. Segui. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24:381–410, 1992.

[44] A. Cutler and D. Norris. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14:113–121, 1988.

[45] A. Cutler and T. Otake. Mora or morpheme? further evidence for language-specific listening. *Journal of Memory and Language*, 33:824–844, 1994.

[46] A. Cutler and T. Otake. Mora or phoneme? further evidence for language specific listening. *Journal of Memory and Language*, 33:824–844, 1994.

[47] C. de Marcken. *Unsupervised Language Acquisition.* PhD thesis, MIT, Cambridge, Massachusetts, 1996.

[48] J. G. de Villiers and P. A. de Villiers. A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2:267–78, 1973.

[49] J. G. de Villiers and P. A. de Villiers. The acquisition of english. In D. I. Slobin, editor, *The Crosslinguistic Study of Language Acquisition, Vol. 1: The data.* Erlbaum, Hillsdale, NJ, 1985.

[50] A. J. DeCasper and W. P. Fifer. Of human bonding: Newborns prefer their mother's voices. *Science*, 208:1174–1176, 1980.

[51] A. J. DeCasper, J. P. Lecanuet, M. C. Busnel, C. Granier-Deferre, and R. Maugeais. Fetal reaction to recurrent maternal speech. *Infant Behaviour and Development*, 17:159–164, 1994.

[52] A. J. DeCasper and M. J. Spence. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behaviour and Development*, 9:133–150, 1986.

[53] C. H. Echols, M. J. Crowhurst, and J. B. Childers. The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36:202–225, 1997.

[54] P. D. Eimas, E. R. Siqueland, P. Jusczyk, and J. Vigorito. Speech perception in infants. *Science*, 171:303–6, 1971.

[55] A. Fernald. Four-month-old infants prefer to listen to motherese. *Infant Behaviour and Development*, 8:181–95, 1985.

[56] A. Fernald. Human maternal vocalization to infants as biologically relevant signals: An evolutionary perspectives. In J. J. Barkow, L. Cosmides, and J. Tooby, editors, *The Adapted Mind: Evolutionary psychology and the generation of culture.* Oxford University Press, Oxford, 1992.

[57] A. Fernald and P. K. Kuhl. Acoustic determinants of infant preference for motherese speech. *Infant Behaviour and Development*, 10:279–293, 1987.

[58] A. Fernald and G. McRoberts. Prosodic bootstrapping: Critical analysis of the argument and the evidence. In J. L. Morgan and A. van Kleeck, editors, *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition*, pages 117–134. Erlbaum, Hillsdale, NJ, 1996.

[59] A. Fernald and T. Simon. Expanded intonation contours in mother's speech to newborns. *Developmental Psychology*, 20:104–113, 1984.

[60] A. Fernald, T. Taeschner, I. Dunn, M. Papousek, B. de Boysson-Bardies, and I. Fukui. A cross-language study of prosodic modification in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16:477–501, 1989.

[61] A. D. Friederici and J. M. I. Wessels. Phonotactic knowledge and its use in infant speech perception. *Perception and Psychology*, 54:287–295, 1993.

[62] O. Garnica. Some prosodic and paralinguistic features of speech to young children. In C. E. Snow and C. A. Ferguson, editors, *Talking to Children: Language input and acquisition*, pages 63–88. Cambridge University Press, Cambridge, 1977.

[63] L. A. Gerken. Child phonology. In M. A. Gernsbacher, editor, *Handbook of Psycholinguistics*. Academic Press, San Diego, CA, 1993.

[64] L. A. Gerken, P. W. Jusczyk, and D. R. Mandel. When prosody fails to cue syntactic structure: nine-month-olds' sensitivity to phonological vs. prosodic phrases. *Cognition*, 51:237–265, 1994.

[65] L. R. Gleitman, H. Gleitman, B. Landau, and E. Manner. Where learning begins: Initial representation for language learning. In F. J. Newmeyer, editor, *Language: Psychological and Biological Aspects*, pages 150–193. Cambridge University Press, Cambridge, England, 1988.

[66] L. R. Gleitman and E. Wanner. The state of the state of the art. In E. Wanner and L. R. Gleitman, editors, *Language Acquisition: The State of the Art*. Cambridge University Press, Cambridge, England, 1982.

[67] S. M. Glenn, C. C. Cunningham, and P. F. Joyce. A study of auditory preference in nonhandicapped infants and infants with down's syndrome. *Child Development*, 52:1303–7, 1981.

[68] J. C. Goodman and H. C. Nusbaum. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. MIT Press, Cambridge, MA, 1994.

[69] J. V. Goodsitt, J. L Morgan, and P. K. Kuhl. Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, 20:229–252, 1993.

[70] J. V. Goodsitt, P. A. Morse, J. N. Ver Hoeve, and N. Cowan. Infant speech recognition in multisyllabic contexts. *Child Development*, 55:903–910, 1984.

[71] P. Gordon. Learnability and feedback: A commentary on bohanon and stanowicz. *Developmental Psychology*, 26:217–20, 1990.

[72] J. Grimshaw and S. Pinker. Positive and negative evidence in language acquisition. *Behavioral and Brain Science*, 12:341–??, 1989.

[73] S. E. Health. *Ways with Words*. Cambridge University Press, Cambridge, England, 1983.

[74] K. Hirsh-Pasek, R. Golinkoff, A. Fletcher, F. DeGaspe Beaubien, and K. Cauley. In the beginning: One word speakers comprehend word order. Paper presented at Boston University Conference on Language Development, Boston, 1985.

[75] K. Hirsh-Pasek, D. G. Kemler Nelson, P. W. Jusczyk, K. W. Cassidy, B. Druss, and L. Kennedy. Clauses are perceptual units for young infants. *Cognition*, 26:269–286, 1987.

[76] E. Hoff-Ginsberg. *Language Development*. Brooks/Cole Publishing Company, Pacific Grove, CA, 1997.

[77] E. A. Hohne, A. P. W. Jusczyk, and M. Newsome. Infants' strategies of speech segmentation: Clues from weak/strong words. Paper presented at the 20th Annual Boston University Conference on Language Acquisition, Boston, MA, 1995.

[78] E. A. Hohne and P. W. Jusczyk. Tow-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, 56:613–623, 1994.

[79] J. Huttenlocher and P. Smiley. Early word meaning: The case of object names. *Cognitive Psychology*, 19:63–89, 1987.

[80] D. Ingram. *First Language Acquisition: Method, Description and Explanation.* Cambridge University Press, Cambridge, 1989.

[81] J. S. Johnson and E. L. Newport. Critical period effect in second language learning: The influence of maturational state on the acquisition of english as a second language. *Cognitive Psychology*, 21:60–99, 1989.

[82] P. W. Jusczyk. A review of speech perception research. In K. Boff, L. Kaufman, and J. Thomas, editors, *Handbook of Perception and Human Performance*, volume 2. Wiley, New York, 1986.

[83] P. W. Jusczyk. Perception of cues to clausal units in native and non-native languages. Paper presented at The Biennial Meeting of the Society for Research in Child Development, Kansas City, 1989.

[84] P. W. Jusczyk. *The Discovery of Spoken Language.* MIT Press, Cambridge, MA, 1997.

[85] P. W. Jusczyk. Constraining the search for structure in the input. *Lingua*, 106:197–218, 1998.

[86] P. W. Jusczyk. How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9):323–328, 1999.

[87] P. W. Jusczyk and R. N. Aslin. Infants' detection of sound patterns of sound patterns of words in fluent speech. *Cognitive Psychology*, 29:1–23, 1995.

[88] P. W. Jusczyk, A. Cutler, and N. Redanz. Preference for the predominant stress patterns of english words. *Child Development*, 64:675–687, 1993.

[89] P. W. Jusczyk and C. Derrah. Representation of speech sounds by young infants. *Developmental Psychology*, 23:648–654, 1987.

[90] P. W. Jusczyk, A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, and A. M. Jusczyk. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32:402–420, 1993.

[91] P. W. Jusczyk, K. Hirsh-Pasek, D. G. Kemler Nelson, L. Kennedy, A. Woodward, and J. Piwoz. Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24:252–293, 1992.

[92] P. W. Jusczyk, E. A. Hohne, and A. Baumann. Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics*, 61:1465–1476, 1999.

[93] P. W. Jusczyk, D. Houston, and M. Newsome. The beginning of word segmentation in english-learning infants. *Cognitive Psychology*, 39:159–207, 1999.

[94] P. W. Jusczyk, P. A. Luce, and J. Charles-Luce. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33:630–645, 1994.

[95] R. G. Karzon. Discrimination of polysyllabic sequences by one- to four-month old infants. *Journal of Experimental Child Psychology*, 39:326–342, 1985.

[96] K. Kaye. Why don't we talk "baby talk" to babies. *Journal of Child Language*, 7:489–507, 1980.

[97] D. G. Kemler Nelson, K. Hirsh-Pasek, P. W. Jusczyk, and K. W. Cassidy. How the prosodic cues in motherese might assist language learning. *Journal of child Language*, 16:55–68, 1989.

[98] Chunyu Kit. *Unsupervised Lexical Learning as Inductive Inference.* PhD thesis, University of Sheffield, 2000.

[99] A. N. Kolmogorov. Three approaches for defining the concept of “information quantity”. *Problem of Information Transmission*, 1:4–7, 1965.

[100] P. K. Kuhl. Speech perception in early infancy. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, and S. R. Silverman, editors, *Hearing and Davis: Essays honoring Hallowell Davis*, pages 265–280. Washington University Press, St. Louis, 1976.

[101] P. K. Kuhl. Perception of speech and sound in early infancy. In P. Salapatek and L. Cohen, editors, *Handbook of Infant Perception*, pages 275–382. Academic Press, New York, 1987.

[102] P. K. Kuhl and J. D. Miller. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190:69–72, 1975.

[103] B. Landau, S. Jones, and L. B. Smith. The importance of shape in early lexical learning. *Cognitive Development*, 3:299–321, 1988.

[104] H. Lane. *The Wild Boy of Aveyron.* Harvard University Press, Cambridge, MA, 1976.

[105] J. P. Lecanuet. Foetal responses to auditory and speech stimuli. In A. Slater, editor, *Perceptual Development: Visual, auditory, and speech perception in infancy*, pages 317–355. Psychology Press, East Sussex, UK, 1998.

[106] J. P. Lecanuet, C. Cranier-Deferre, and M. C. Busnel. Human fetal auditory perception. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, and W. P. Smotherman, editors, *Fetal Development: A psychobiological perspective*, pages 239–262. Lawrence Erlbaum, Hillsdale, NJ, 1995.

[107] J. P. Lecanuet, C. Granier-Deferre, A. J. DeCasper, R. Maugeais, A. J. Andrieu, and M. C. Busnel. Perception et discrimination foetale de stimuli langagier, mise en évidence à partir de la réactivitaté cardiaque. résultant prélimimaires. *Compte-Rendus de l’Académie des Sciences de Paris*, 305:279–303, 1987.

[108] J. P. Lecanuet, C. Granier-Deferre, A. Y. Jacquet, and M. C. Busnel. Decelerative cardiac responsiveness to acoustical stimulation in the near term foetus. *Quarterly Journal of Experimental Psychology*, 44B:279–303, 1992.

[109] J. P. Lecanuet, C. Granier-Deferre, A. Y. Jacquet, I. Capponi, and L. Ledru. Prenatal discrimination of a male and female voice uttering the same sentence. *Early Development and Parenting*, 2:217–228, 1993.

[110] I. Lehiste. *An Acoustic-phonetic Study of internal Open Juncture.* S. Karger, New York, 1960.

[111] I. Lehiste. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51:2018–2024, 1972.

[112] M. Li and P. M. B. Vitányi. *Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, New York, 1993. Second edition, 1997.

[113] K. Malmkjær. Language acquisition. In K. Malmkjær, editor, *The Linguistics Encyclopedia*, pages 239–251. Routledge, London, 1991.

[114] D. R. Mandel, P. W. Jusczyk, and D. G. Kemler Nelson. Does sentential prosody help infants organize and remember speech information. *Cognition*, 53:155–180, 1994.

[115] D. R. Mandel, P. W. Jusczyk, and D. B. Pisoni. Infants' recognition of the sound patterns of their own names. *Psycholinguistic Science*, 6:315–318, 1995.

[116] G. F. Marcus. Negative evidence in language acquisition. *Cognition*, 46:53–85, 1993.

[117] S. L. Mattys and P. W. Jusczyk. Phonotactic cues for segmentation of fluent speech by infants. Manuscript, Dept. of Psychological and Cognitive Science, Johns Hopkins University, 2000.

[118] S. L. Mattys, P. W. Jusczyk, P. A. Luce, and J. L. Morgan. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38:465–494, 1999.

[119] J. L. McClelland and J. L. Elman. The TRACE model for speech perception. *Cognitive Psychology*, 18:1–86, 1986.

[120] J. M. McQueen. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39:21–46, 1998.

[121] J. M. McQueen, D. Norris, and A. Cutler. Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:621–638, 1994.

[122] J. Mehler, P. W. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison. A precursor of language acquisition in young infants. *Cognition*, 29:143–78, 1988.

[123] S. Messer. Implicit phonology in children. *Journal of Verbal Learning and Verbal Behavior*, 6:609–613, 1967.

[124] J. D. Miller, C. C. Wier, R. E. Pastore, W. J. Kelley, and R. J. Dooling. Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorial perception. *Journal of the Acoustical Society of American*, 60:410–417, 1976.

[125] J. L. Miller and P. D. Eimas. Observations on speech perception, its development, and the search for a mechanism. In J. C. Goodman and H. C. Nusbaum, editors, *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, pages 37–56. MIT Press, Cambridge, MA, 1994.

[126] J. L. Morgan. *From Simple Input to Complex Syntax.* MIT Press, Cambridge, MA, 1986.

[127] J. L Morgan. Converging measures of speech segmentation in prelingual infatns. *Infant Behavior and Development*, 17:389–403, 1994.

[128] J. L. Morgan. A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, 35:666–688, 1996.

[129] J. L. Morgan and J. R. Saffran. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66:911–936, 1995.

[130] J. L. Morgan and L. L. Travis. Limits on negative information on language learning. *Journal of Child Language*, 16:531–52, 1989.

[131] P. A. Morse. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 14:477–92, 1972.

[132] J. Myers, P. W. Jusczyk, D. G. Kemler Nelson, J. Charles-Luce, A. Woodward, and K. Hirsh-Pasek. Infants' sensitivity to word boundary in fluent speech. *Journal of Child Development*, 23:1–20, 1996.

[133] L. H. Nakatani and K. D. Dukes. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62:714–719, 1977.

[134] T. Nazzi, J. Bertoncini, and J. Mehler. Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24:756–766, 1998.

[135] K. Nelson. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 38, 1973.

[136] E. L. Newport. Contrasting concept of critical period for language. In S. Carey and R. Gelman, editors, *The Epigenesis of Mind: Essays on biology and cognition*. Erlbaum, Hillsdale, NJ, 1991.

[137] E. L. Newport, H. R. Gleitman, and L. R. Gleitman. Mother i'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow and C. A. Ferguson, editors, *Talking to Children: Language input and acquisition*, pages 109–150. Cambridge University Press, Cambridge, 1977.

[138] E. L. Newport and R. P. Meicer. The acquisition of American Sign Language. In D. I. Slobin, editor, *The Crosslinguistic Study of Language Acquisition, Vol. 1: The data*. Erlbaum, Hillsdale, NJ, 1985.

[139] M. Newsome and P. W. Jusczyk. Do infants use stress as a cue for segmenting fluent speech? In D. MacLaughlin and S. gMcEwen, editor, *Proceedings of the 19th Annual Boston University Conference on Language Development* (Vol. 2), pages 415–426, Somerville, MA, 1995. Cascadilla.

[140] D. Norris, J. M. McQueen, and A. Cutler. Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:1209–1228, 1995.

[141] D. Norris, J. M. McQueen, and A. Cutler. The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34:191–243, 1997.

[142] D. G. Norris. Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52:189–234, 1994.

[143] E. Ochs. Talking to children in western samoa. *Language in Society*, 11:77–104, 1982.

[144] D. C. Olivier. *Stochastic Grammars and Language Acquisition Mechanisms*. PhD thesis, Harvard University, Cambridge, MA, 1968.

[145] T. Otake, G. Hatano, and A. Cutler. Mora or syllable? speech segmentation in japanese. *Journal of Memory and Language*, 32:258–278, 1993.

[146] T. Otake, G. Hatano, A. Cutler, and J. Mehler. Mora or phoneme? further evidence for language specific listening. *Journal of Memory and Language*, 32:258–278, 1993.

[147] A. Peters. *The Units of Language Acquisition*. Cambridge University Press, Cambridge, England, 1983.

[148] L. A. Petitto. *From Gesture to Symbol: The relationship between form and meaning in the acquisition of personal pronoun in American Sign Language*. PhD thesis, Harvard University, Boston, MA, 1984.

[149] L. A. Petitto. On the autonomy of language and gesture: Evidence from the acquisition of american sign language. *Cognition*, 27(1):1–52, 1987.

[150] L. A. Petitto. "language" in the pre-linguistic child. In K. S. Kesel, editor, *Development of Language and Language Researchers: Essays in honor of Roger Brown*, pages 187–221. Erlbaum, Hillsdale, NJ, 1988.

[151] L. A. Petitto. Modularity and constraints in early lexical acquisition: Evidence from children's early language and gesture. In P. Bloom, editor, *Language Acquisition: Core readings*, pages 95–126. MIT Press, Cambridge, MA, 1994.

[152] L. A. Petitto and P. F. Marentette. Babbling in the manual mode: Evidence foe the ontogeny of language. *Science*, 251:1493–6, 1991.

[153] S. Pinker. *Language Learnability and Language Development.* Harvard University Press, Cambridge, MA, 1984.

[154] S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Science*, 13:707–78, 1987.

[155] M. A. Pitt and J. M. McQueen. In compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39:347–370, 1998.

[156] J. Rissanen. Modelling by shortest data description. *Automatica*, 14:465–471, 1978.

[157] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, N.J., 1989.

[158] R. Rymer. *Genie: An abused child's flight from silence.* HarperCollins, New York, 1993.

[159] J. R. Saffran, R. N. Aslin, and E. L. Newport. Statistical learning by 8-month-old infants. *Science*, 274:1926–8, December 1996.

[160] B. B. Schieffelin. Getting it together: An ethnographic approach to the study of the development of communicative competence. In E. Ochs and B. B. Schieffelin, editors, *Developmental Pragmatics*, pages 73–110. Academic Press, New York, 1979.

[161] B. B. Schieffelin. The acquisition of kaluni. In D. I. Slobin, editor, *The Crosslinguistic Study of Language Acquisition: Vol. 1. The Data*, pages 525–594. Erlbaum, Hillsdale, NJ, 1985.

[162] B. B. Schieffelin and E. Ochs. Language socialization. *Annual Review of Anthropology*, 15:163–191, 1986.

[163] R. Schreuder and R. H. Baayen. Prefixing stripping re-visited. *Journal of Memory and Language*, 33:357–375, 1994.

[164] R. Shi, J. F. Werker, and J. L. Morgan. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72:B11–B21, 1999.

[165] N. N. Soja. Inferences about the meaning of nouns: The relationship between perception and syntax. *Cognitive Development*, 7:29–45, 1992.

[166] R. J. Solomonoff. A formal theory of inductive inference, part 1 and 2. *Information Control*, 7:1–22, 224–256, 1964.

[167] D. N. Stern, S. Spieker, R. K. Barnett, and K. MacKain. The prosody of maternal speech: Infant age and context-related changes. *Journal of Child Language*, 10:1–15, 1983.

[168] S. E. Trehub. Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9:91–96, 1973.

[169] S. E. Trehub. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47:466–472, 1976.

[170] N. Umeda and C. H. Coker. Allophonic variation in American English. *Journal of Phonetics*, 2:1–5, 1974.

[171] M. S. Vitevitvch and P. A. Luce. Probabilistic phonotatics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40:374–408, 1999.

[172] M. S. Vitevitvch, P. A. Luce, J. Charles-Luce, and D. Kemmerer. Phonotatics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40:47–62, 1997.

[173] J. Vroomen and B. de Gelder. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21:98–108, 1995.

[174] J. Vroomen, M. van Zon, and B. de Gelder. Cues to speech segmentation: Evidence from juncture misperception and word spotting. *Memory and Cognition*, 24:744–755, 1996.

[175] J. F. Werker, J. H. V Gilbert, K. Humphrey, and R. C. Tees. Developmental aspects of cross-language speech perception. *Child Development*, 52:349–355, 1981.

[176] J. F. Werker and R. C. Tees. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7:49–63, 1984.

[177] C. C. Wood. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of American*, 60:1381–1389, 1976.

[178] G. Yule. *The Study of Language: an Introduction*. Cambridge University Press, Cambridge, 1985.

[179] G. K. Zipf. *The Psycho-Biology of Language*. Houghton Mifflin, 1935.

[180] G. K. Zipf. *Human Behaviour and the Principle of Least Effort*. Hafner, New York, 1949.