Infant word segmentation: a basic review

Morgan Sonderegger*

4/23/08

1 Introduction

The abstract problem of speech segmentation can be formulated computationally: given strings ($s_i \in \Sigma^*$) over some alphabet (Σ) and the knowledge that each string (s_i) is a concatenation of words from some set ($W \subseteq \Sigma^*$), how does the learner determine what the words (W) are and segment new strings? Even assuming that the alphabet (Σ) is known (Σ can be {syllables}, {segments}, ...), this is a daunting problem.

In the context of natural language, this abstract problem intuitively seems overly difficult, since written words are separated by spaces and spoken words seem separated by pauses. However, as can be seen in a spectrogram of spoken English, pauses at word boundaries are in fact no longer than between arbitrary syllables; indeed phoneticians have found *no* systematic cues to word boundaries in speech [17, 19, 18]. So the segmentation problem faced by humans is at least as hard as that described above, yet infants routinely solve it early in language acquisition. Segmentation is easy enough for humans that many written languages do not include spaces.[1]

How do infants learn to segment their native language(s)? The consensus of most experimenters is that infants gradually integrate a range of cues, each of which is insufficient in isolation, to segment in progressively greater detail. This general statement admits a range of hypotheses, since for each potential cue one can ask five questions:

1. Is the cue reliably present in the infant's input?
2. Can infants use the cue to segment (in a controlled test)?
3. How does the cue interact with other cues?
4. Does the cue occur in infant-directed input cross-linguistically?
5. Are the mechanisms used by infants to exploit the cue specific to language, or domain-general?

(5) (in a more general form) has perhaps seen the most debate, little of which is directly relevant to how the segmentation problem is solved. We stick to (1)-(4) in this review to focus on the central question: *how are multiple sources of information in the infant's input integrated together to perform segmentation, and how does this process develop*?

There is a large literature on word segmentation by infants, and because the topic is important to some larger theoretical issues (outlined below) is is often written about in the context of larger debates. It can therefore be difficult to get a handle on exactly what is known about how infants begin to segment words (questions (1)-(4)). This review attempts a rough survey of work on word segmentation by infants, with some emphasis on placing well-known results in the broader experimental context. The aim is not to be comprehensive, but to provide a balanced sample of the substantial literature and cover all major results. We first discuss background methods and concepts.

*Thanks to Terry Regier for his continued patience.

[1] Including Japanese, Khmer, Latin, and Old Egyptian, indicating that not using spaces does not seem to be restricted to certain language families or geographical areas. Spacing was only introduced in European languages in medieval times. [81]

2 Preliminaries

2.1 The Headturn Preference Procedure

Most studies discussed below use a variant of the Headturn Preference Procedure (HPP: [63]) to assess infants' reactions to stimuli. Essentially, a flashing light causes the subject to turn his head towards a speaker, a test stimulus begins playing from it, and the time it takes for him to become bored and turn his head away is measured. Each subject is exposed to control and test items, and the null hypothesis – that test and control orientation times are the same – is tested over a group of subjects. The surprising reliability of this simple procedure is described in [63].

The one complication of HPP is that depending on the experimental task, infants sometimes display a "novelty preference" (longer orientation times for novel items than for familiar ones) and sometimes a "familiarity preference" (vice versa), roughly depending on task difficulty [50]. This complicates comparing HPP results across multiple experiments, since if infants are habituated to A, then pay attention longer to A than B in Expt. 1 and longer to A than C in Expt. 2, they could prefer A in Expt. 1 (familiarity) but C in Expt. 2 (novelty). Since experimenters are aware of this issue and include extra controls and conditions to deal with it, it is not discussed further here.

2.2 The unit of segmentation

To study segmentation of the speech stream, it is necessary to know what units infants are segmenting over. The intuitive options are segments and syllables, and many studies on how infants of different ages process and store linguistic detail address which one (or neither) infants use over time. Although the issue is not resolved, there is general agreement that syllables are the primary unit of analysis during much of the first year, when segmentation abilities develop [30, 55, 72]. Evidence comes from studies showing that young ($<$ 4 months) infants can recognize when words share syllables, but not individual segments [5, 31, 54, 60], that infants are aware of segments as a unit later than syllables [65], and that gains in the ability to generalize over classes of segments develop gradually over the first year [33, 41, 42, 56]. However, as discussed in 3.3.3, segmentally-based cues for segmentation come into play later in the first year, and different developmental trajectories are observed for syllable types differing by segmental content.

2.3 Transitional probabilities

The hypothesis ($\mathcal{H}$) which sparked interest in statistical learning for word segmentation dates back at least to a 1955 article by Zellig Harris [39], but lay mostly unexamined for decades: in a stream $a_1 a_2 \cdots a_n$ of linguistic units which must be segmented into words, the higher the conditional probability $P(a_i \,|\, a_{i-1})$, the more likely a_i and a_{i+1} are to belong to the same word.[2] Put otherwise, word boundaries often occur at relative minima of unit-unit conditional probability.[3] This seems intuitively plausible because words can be used in different contexts, meaning syllables which occur word-finally should have low conditional probability with any following syllable. The example often given is that an infant is likely to have heard many *pretty+noun* examples but few words ending in *pre-* or beginning in *-ty*.

However, $\mathcal{H}$ is an empirical claim that must be validated for a given language at a given level of representation (i.e. segments or syllables). As discussed below (3.3.4) in the context of the "trochaic bias" of West Germanic languages (English, German, Dutch), intuition can be misleading, especially when considering child-directed speech. And since most studies showing sensitivity to asymmetries in TPs use artificial languages, such results are only relevant to natural language if similar patterns occur there.

For American English, $\mathcal{H}$ has recently been investigated and confirmed at the segmental level by S. Hockema [43] and at the syllabic level by D. Swingley [95], but with important differences between the two.

[2] Harris worked in terms of "successor frequencies/counts" (SF), not probabilities, but the concepts are equivalent. His idea was to discover morpheme rather than word boundaries; ironically, this idea turned out to be much more effective for the latter than the former.

[3] The term "transitional probability" (TP) is usually used in SL and other segmentation literature instead of "conditional probability"; for consistency TP is used here.

Hockema investigated phoneme-phoneme transitions in the CHILDES corpora [66], estimating phonemic transcription using a pronouncing dictionary, and found that the distribution of P_{wb} (the fraction of examples seen at a word boundary, for a given phoneme pair) over attested phoneme-phoneme pairs is strongly bimodal: 52% of pairs have $P_{wb} < 0.02$ or $P_{wb} > 0.98$, and 79% have $P_{wb} < 0.2$ or $P_{wb} > 0.8$. The bimodal pattern does not change much when pairs are scaled by frequency of occurrence, or (more surprisingly) when the input corpus is generated randomly from an English dictionary using Kucera-Francis word frequencies.

Swingley implemented a syllable-based segmentation algorithm on corpora of child-directed speech for English [64] and Dutch [103]. Although a raw distribution of P_{wb} for syllable-syllable pairs is not given, the algorithm outputs something very similar by grouping two syllables together if their mutual information (MI, closely related to TP[4]) and individual frequencies, calculated over the corpus, are above a threshold (expressed as a percentile over MI+frequency scores for all syllable-syllable pairs). Graphs of accuracy (true positives/(true positives+false positives)) versus the threshold value show that accuracy for hypothesized bisyllabic words start very low, but go up sharply as the threshold is increased past 75%, to > 0.75, in both English and Dutch. The same pattern occurs for trisyllabic words in Dutch, but not in English. As the threshold value increases the number of hypothesized words goes down, so completeness suffers as accuracy improves. Put otherwise, $\mathcal{H}$ holds *if* a high enough threshold value is used, and at least in English only for bisyllabic words. Swingley also found that these results held up

1. when some noise was introduced into the location of syllable boundaries, important because infants are concurrently learning phonotactics.

2. when small time-slices of the corpus were used, important because infants do not remember all words ever spoken to them.

In sum, the distributional properties of the input for English support $\mathcal{H}$ over segments, and support $\mathcal{H}$ over syllables if some conditions are assumed on the learner. Differences in the results for even typologically-similar English and Dutch suggest that future cross-linguistic study is needed, and both studies suggest that $\mathcal{H}$ cannot be taken for granted.

2.4 Prosody, statistics: background

Much work in infant word segmentation has focused on prosodic cues or distributional information (TPs between segments or syllables) in the input. These are the primary examples of what can be termed "bracketing cues" and "clustering cues", discussed below. Although there is no a priori reason the two should be mutually exclusive, work on segmentation by infants is often summarized as "prosody versus statistics". We provide some experimental background for prosodic and statistical cues.

Prosody

Work from the late 1980s on demonstrated that adult English speakers segment novel speech into trochaic feet [21, 22, 23, 27], in general postulating word boundaries at strong syllables. In a study of British conversational speech, notable as one of the first corpus-based studies of pronunciation, Cutler and Carter [24] found that approximately 90% of content words began with a strong syllable and 75% of non-initial syllables were weak, making the simple strategy "postulate a new word at each S" surprisingly plausible. This led to the more general Metrical Segmentation Strategy (MSS) hypothesis, that listeners segment based on the "rhythm" or "metrical structure" of their language.[5] Cross-linguistic support came from studies on syllable-timed French [25, 26] and mora-timed Japanese [82] indicating that speakers of these languages segment consistently with rhythms which are syllable/mora-based rather than foot-based.[6] Support for language-

[4]For a sequence AB, TP=$P(B \mid A) = P(AB)/P(A)$, while MI=$\log_2[P(AB)/(P(A)P(B))]$, so MI = $\log_2(\text{TP}/P(B))$. MI is basically symmetrized TP.

[5]More formally, a language's rhythmic structure is based on its prosodic tier which is most periodic, but the question of exactly what (acoustically) talking about a language's rhythm means remains largely open [84]

[6]The adequacy of the "X-timed language" description to describe rhythm has been the subject of much debate; for example see [105] on "mora-timed" Japanese. Here it only matters that rhythm in different languages emphasizes different prosodic constituents, which is not disputed.

specific MSS in adult processing naturally led to work on its use by infants, discussed below.

Statistics
Few experiments before Saffran, Aslin, and Newport's landmark 1995 study (SAN) [88] tested the hypothesis that humans can use distributional information for segmentation.[7] A 1970 study [40] by Hayes and Clark exposed adult listeners to a 45-minute stream of concatenated words (with no pauses, etc.) from an "artificial speech analog" consisting of words made up of sounds alternately from two classes of noises. These were analogous to consonants and vowels, which tend to alternate. After this exposure, listeners could differentiate between sequences of words with pauses between words and sequences with pauses within words.

Goodsitt, Morgan, and Kuhl (1993) (GMK: [37]) were the first to examine whether infants can use distributional information in speech. Eight-month-olds were trained on sets of trisyllabic words from small artificial languages over an alphabet of four syllables. GMK's experiments are difficult to describe briefly, but essentially showed that infants' ability to correctly decide a novel word's membership in an artificial language she has been trained on goes up significantly when the language includes an "invariant ordering" of two syllables: this is the case for language {ABC, CAB, DAB, ABD} but not for language {ABC, CBA, DAB, BAD}. Put otherwise, infants can store and use *asymmetric* transitional probability information.

GMK distinguishes between two strategies for segmentation infants could use: "bracketing strategies" (such as MSS) which use prosodic cues to find word boundaries, and "clustering strategies" (such as GMK's findings) which find groups of syllables separated by high transitional probabilities. Generalized somewhat, the bracketing/clustering distinction is useful for thinking about all proposed cues for word segmentation: bracketing cues signal potential word boundaries, and clustering cues signal potential words or groups of words (e.g. "stickiness" between syllables).

2.5 Computational Models

Among all aspects of acquisition, word segmentation is perhaps most easily stated as a computational problem. There is a long lineage of computational models of word segmentation in a variety of frameworks, and describing them is beyond the scope of this paper. Brent [8] provides a concise review of the main models up to 1999 [3, 9, 7, 11, 13, 28, 32, 80, 85, 89, 106], and more recent work includes [4, 34, 95].

3 Main

Interest in infant word segmentation intensified following two elegant studies: SAN (1996) and Jusczyk and Aslin (1995) [53] (JA). Because their methods are used in most later work, both are discussed in some detail.

Jusczyk and Aslin (1995)
JA familiarized 7.5-month-old English-learning infants with two monosyllabic English words, then exposed them to sentences containing and not containing one repetition of a familiar word. Using a modified HPP procedure, JA found that infants showed a significant novelty preference for sentences containing familiar words. Subsequent experiments found that 6-month-olds do not show this preference, and neither do 7.5-month-olds if the test sentences include familiar words with the initial segment changed by one feature. However, since infants do not first hear most words in isolation, JA also reversed the paradigm: 7.5-month-olds were played a series of sentences containing one of two target words, then tested on individual words, including target words; again the infants showed a novelty preference for target words, demonstrating segmentation from fluent speech.

In JA's experiments, all speech was "in a lively voice, as if naming the object for an infant" (6). Infants heard training words spoken 15 (different) times, and in the last experiment heard at most 12 sentences containing the test word, with total duration of approximately 45 seconds.

That rudimentary segmentation develops between 6-7.5 months is remarkable given that (1) Basic word comprehension begins only at about 9 months, and takes until 15 months to fully develop. (2) Infants do

[7] Using human subjects, as opposed to computational models; on these see 2.5

not begin to become attuned to sub-prosodic details of their native language until 6-9 months. This means infants can segment words long before they have real awareness *of* words as a meaningful unit. Since infants must be treating the words simply as familiar clusters of sound, it is not surprising that they no longer do so when the initial segment is changed.

Saffran, Aslin, and Newport (1996)
SAN tested whether eight-month-old infants could segment based on distributional cues alone at the syllable level. Artificial languages such as these were used:

L_1: *tupiro*, *golabu*, *bidaku*, *padoti*

L_2: *dapiku*, *tilado*, *burobi*, *pagotu*

Note that each syllable occurs only once per language, so within-word syllable sequences have TP=1, while between-word syllable sequences have TP=$\frac{1}{3}$.

Subjects were exposed to a two-minute stream of synthesized speech from one language with no prosody and no pauses between words, such as (L_1) *tupirogolabubidakupadotibidakugolabutupiro...* They were then tested using HPP on test stimuli *tupiro*, *golabu*, *dapiku*, *tilado*, consisting of two words and two non-words made up of syllables present in the input but ordered differently. There was a significant novelty preference for non-words, indicating subjects tracked serial order of the syllables.

Another set of subjects was played speech streams, but this time tested on stimuli (using L_1 again) *tupiro*, *golabu*, *dakupa*, *dotibi* consisting of two words and two part-words, meaning a trisyllabic sequences in the speech stream which crossed word boundaries. There was a significant novelty preference for part-words, indicating that subjects must have been tracking TPs between adjacent syllables (since no other information to differentiate between words and part-words was available).

These findings are as striking as JA's: if it is impressive that 7.5-month-old infants can segment a recurring monosyllable from a speech stream, it is equally remarkable that 8-month-olds can not only segment recurring trisyllables, but differentiate between them based on distributional statistics alone.

SAN has had an enormous impact both among researchers who study language and the wider scientific community. To see this, searching the Scopus and ISI citation databases[8] shows that SAN is the most-cited work on language acquisition of the past 25 years, as well as the fifth-most-cited of *all* research work on human language of the past 25 years, and among research by linguists is second only to Chomsky's *Minimalist Program*. The reasons for SAN's impact are discussed following a comparison with JA, which has been arguably as influential in language acquisition, but not beyond.

3.1 JA, SAN and statistical learning

JA and SAN both demonstrate the use of clustering cues by infants, but with very different emphases. JA use natural speech, make essentially no theoretical points, and say their contribution was to show that 7.5-month-old infants "have at least some rudimentary capacity to recognize words in fluent speech contexts", but caution that only monosyllables were used, etc. The amount of speech infants were exposed to (maximum 45 seconds) is hardly mentioned. In contrast, SAN exclude non-statistical cues by using synthesized speech, but complicate the task of observing distributional information by using trisyllables. Infants are characterized as able to accomplish "a fundamental task of language acquisition [word segmentation] based solely on the statistical relationships between speech sounds", and the fact that infants had "only 2 minutes" of exposure is mentioned several times. These differences can be ascribed in part to different audiences: JA published in a psychology journal, SAN in *Science*.

But the impact of SAN is best explained by its theoretical claims, summed up in the conclusion:

> Our results raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language. It remains unclear whether the statistical learning we observed is

[8] Accessed 4/6/08.

> indicative of a mechanism specific to language acquisition or of a general learning mechanism... Regardless, the existence of computational abilities that extract structure so rapidly suggests that it is premature to assert a priori how much of the striking knowledge base of human infants is primarily a result of experience-independent mechanisms. In particular, some aspects of early development may turn out to be best characterized as resulting from innately-biased statistical learning mechanisms rather than innate knowledge.

The issues raised are exactly those which have fueled subsequent work and debate, and indicate why SAN has been cited so often: it was crucial in fueling (starting?) a larger debate over three questions:

1. To what extent are the mechanisms underlying human language domain-general vs. domain-specific?
2. To what extent do they use/build statistical generalizations vs. discrete rules and representations
3. How much domain-specific knowledge are infants born with?

Most importantly (and underlying all three questions), SAN explicitly question the primacy of innate linguistic knowledge ("experience-independent mechanisms"), which in a footnote is identified as referring to the "poverty of the stimulus" argument, the central dogma of generative linguistics. It is clear that the authors meant to make a bold theoretical statement and start the resulting debate.

Work on statistical learning (SL) has mushroomed since SAN. We will only discuss SL studies relevant to infant word segmentation, but note that there is now a significant literature on SL in word segmentation by adults and older children, applied to higher levels of linguistic structure, in non-linguistic tasks, and by animals. Reviews have been written on infant artificial-grammar studies [36], statistical language learning [35, 87], statistical learning in linguistic and non-linguistic domains [2], and the general idea of domain-general statistical mechanisms [91]. A significant literature from several corners of linguistics arguing against the importance of SL for language acquisition has also developed, for example [12, 34, 73, 67] (and refs). The scope of demonstrations of SL convincingly show that it is a domain-general skill that may be basic to animal cognition; how important it is for language learning in particular is a different question.

3.2 Checking and perturbing SAN

Given the novel results and provocative theoretical claims in SAN, a number of experimental studies have checked its assumptions and found them sound, but have also shown the limitations of the statistical-learning mechanism used in isolation.

3.2.1 Checks

A potential alternative explanation for SAN's results is that subjects used frequency information rather than TPs to distinguish between words and part-words.[9] If infants were using frequencies rather than TPs, SAN's finding would be reduced to a less-surprising one, that infants can distinguish between more and less-frequent sequences. However, a later study by the same authors [1] showed that infants still prefer part-words when the experiment is replicated with frequency controlled for, suggesting that TPs are indeed being used.

Given that SAN used an artificial language made of synthesized speech, one could ask whether infants actually treat the words which they segment out of the speech stream as linguistic objects. Saffran [86] examined whether the "output of statistical learning" is linguistic. Eight-month-old infants were habituated to a (synthesized) speech stream from SAN. One group of infants was then tested using HPP on words vs. part-words in utterance-final English contexts ("I like my *pabiku*"), and another group on words vs. part-word in nonsense contexts ("Zy fike ny *pabiku*") which rhymed with English contexts, both read by an English speaker. The English group showed a familiarity preference for words over part-words while in

[9] Assuming the speech streams used in SAN are made up of four words randomly concatenated, each word occurs (on average) once per 16 syllables while each part-word occurs 1/3 as often.

the Nonsense group there was no difference.[10] Other experiments provided additional evidence that the trisyllables infants segment are being treated as linguistic objects. (For a similar result for JA, see 3.5.1)

A further potential objection to SAN is that its use of synthesized speech is somehow problematic; this factor was addressed by Johnson and Jusczyk [52] who replicated SAN's results using syllables spoken in isolation (to avoid coarticulation) and at flat pitch.

3.2.2 Perturbations

A number of studies have used modified versions of SAN's experiments from the standpoint of checking how powerful statistical learning is as a segmentation tool.

Attention
Linguistic statistical learning has been claimed to occur even during non-linguistic distractor tasks, on the basis of a study where adults were asked to draw while listening to SAN-like stimuli [90]. Toro, Sinnett and Soto-Faraco [99] checked the hypothesis "that word segmentation based on statistical regularities occurs without the need of attenton". In a series of experiments, groups of adult subjects listened to speech streams from SAN while (control group) performing no concurrent task or (test groups) performing concurrent auditory, visual, or stimulus-related tasks; subjects were then tested for word vs. non-word recognition. All concurrent-task groups had worse segmentation performance than control, though how much worse depended on the task; about half of concurrent-task groups segmented at chance level. This result does not directly apply to word segmentation by infants, but does show that using TPs for segmentation likely depends on attention.

Word length
The language used in SAN is clearly much simpler than natural language input, which contains a large number of words of variable length, and it seems reasonable to ask how robust SAN's findings are to changes in the artificial language. Johnson and Jusczyk [51] addressed this question by replicating SAN as closely as possible using modified artificial languages. Johnson and Jusczyk's languages contained two three-syllable, one four-syllable, and one two-syllable word, in contrast to the 4 three-syllable words in SAN's languages; eight-month-old infants were otherwise tested as in SAN for discrimination between words and part-words. Subjects did not show a preference between words and part-words, a null result which held up when subjects were only tested on trisyllabic words and part-words, removing the possible confound of variable word length in the testing phase. Although a basis for comparison (for example checking whether older children or adults can segment the artificial languages used) is lacking, this result is perhaps the most elegant demonstration of limitations on the statistical-learning capabilities found in SAN, especially when used in isolation or on natural language.

Cross-linguistic segmentation
Also on this point are two studies [78, 101] from P. Jusczyk's lab on segmentation of Chinese by English-learning and Chinese-learning infants. These studies were cut short by Jusczyk's untimely death, but seem to have solid preliminary findings. Briefly, if SAN's finding that segmentation of an artificial language is possible after brief exposure generalizes to natural languages without restrictions, one would expect that English-learning infants would begin to be able to segment Chinese given appropriately scaled-up exposure. However, English-learning infants who were exposed to up to 5 hours of child-directed speech in Chinese over a period of several days did not show a preference for words or part-words in a Chinese-language version of SAN, while Chinese-learning infants did.[11] A similar null result for English-language infants but (preliminary) positive result for Chinese-language infants was also obtained for a Chinese-language version of JA.

[10] Note that infants prefer words in this experiment, but part-words in SAN. This familiarity vs. novelty effect (see 2.1) is discussed at some length, but the important point is that infants are treating words and part-words differently in both experiments.

[11] The latter fact has only been tentatively established due to the difficulty of finding monolingual Chinese-learning infants. It is important as a baseline showing that Chinese is segmentable by (Chinese-learning) 8-month-olds.

Taken together, these studies suggest that SL capabilities do not generalize to arbitrary linguistic input.

Discussion
The more general point is that caution is needed before drawing conclusions for natural language-learning from demonstrations of statistical learning in artificial languages. While statistical learning is an interesting and potentially powerful mechanism, especially given its connection to learning in other cognitive domains and by other species, in infant word segmentation in particular SL abilities (in isolation) are fragile and not robust when the input is made more similar to natural language.

However, these limitations on the role of SL in word segmentation should not come as a surprise, and do not show that it is not an essential part of how infants segment. The best-established fact about segmentation, especially early segmentation (6-9 months), is that all proposed cues are fragile and frequently violated in the input. From this perspective, SL is simply another cue that infants learn to integrate for segmentation, albeit a potentially crucial one.

3.3 Non-statistical cues for segmentation

Research has identified several types of non-statistical cues for infant word segmentation, all of which, like statistical learning, are not fully reliable used in isolation. Because of the volume of literature on this topic, studies are reviewed more briefly.

3.3.1 Prosodic cues

As discussed above (2.4), adults seem to predominantly use segmentation strategies specific to the rhythmic structure of their native languages. Since infants are sensitive to the prosodic characteristics of their native language from birth [75], several studies have addressed the extent to which infants use prosodic cues to segment and how these language-specific skills develop.

English
Jusczyk, Houston, and Newsome (JHN) [59] carried out 15 experiments, all variants of JA, on segmentation of bisyllabic words by English-learning infants. They found that (1) 7.5-month-old infants can segment SW words from fluent speech and are segmenting the (SW) chunk rather than the S syllable (2) 7.5-month-old infants cannot segment WS words from fluent speech, but can segment the S syllable. When a WS word occurs before a weak syllable, they segment the (false) SW word (such as *taris* from *guitar is*) (3)10.5-month-old infants can segment WS words from fluent speech, are segmenting the (WS) chunk rather than the S syllable, and do not missegment when a WS word occurs before a weak syllable.

These findings build on an earlier study by Echols, Crowhurst, and Childers [29] where infants and adults were played stimuli with pauses inserted at syllable boundaries. They found that 9-month-olds prefer stimuli where pauses do not interrupt SW sequences, 7-month-olds show no preference for pause location, and 9-month-olds prefer SW sequences previously seen in a four-syllable string to novel SW sequences, but show no preference when WS sequences are used.

This study and JHN suggest the development of a "trochaic bias" in English-learning infants between 7 and 9 months, and suggest the hypothesis that a trochaic MSS is used as a first pass at segmentation ("prosodic bootstrapping"). Infants would then somehow refine their segmentation strategy to pick out less-common stress patterns by 10.5 months. However, a trochaic MSS does not account for infants' abilities by 10.5 months, and is not by iteself sufficient for segmentation. We return below to the question of whether it can at least form a first pass at segmentation.

JHN's results also suggest that the foot is the relevant unit of segmentation for English-learning infants rather than individual syllables. Assuming this is because their ambient language is stress-timed, do learners of a syllable-timed language behave differently?

French
Nazzi et. al. [77] addressed this question by using similar methodology to JHN to test segmentation of WS

words versus final syllables in French infants at 8, 12, and 16 months. (Stress is predominantly word-final in French.[12]) 8-month-olds segmented neither words nor final syllables, 12-month-olds segmented final syllables but not words, and 16-month-olds segmented words but not final syllables. These results suggest a plausible interpretation: learners of syllable-timed French can first segment only salient syllables, then bootstrap to whole words. In doing so they must drop the preference for individual syllables, which would otherwise lead to missegmentation of most word endings as whole words.

A recent bilingual study by Höhle et. al. using French and German infants [44] showed that French-learning infants treat non-native prosodic units analogously to English-learning infants. French six-month-olds could distinguish between SW and WS German stimuli after 1 minute of exposure to SW stimuli, but in a different experiment did not differentiate (measured by looking time) between individually presented SW and WS stimuli (at an age when German-learning infants did). These infants can thus differentiate between SW and WS stimuli, but do not *prefer* one because syllables are the relevant unit in their language. Analogously, JHN (above) found that infants could differentiate between W and S syllables, but did not use the contrast in segmentation.

3.3.2 Acoustic cues

The problem of speech segmentation stems from the intuitively surprising fact that there is no systematic acoustic marking of word boundaries [17, 18, 19]. However, there are systematic cues to *other* linguistic structures which usually occur at word boundaries. Some experiments have tested whether infants can use these cues for segmentation.

Prosodic edges
Best documented is lengthening of syllables which fall at the edges of prosodic constituents (see [62]). The effect increases going up the prosodic hierarchy from feet to utterance and final effects are generally stronger than initial ones.

In infant-directed speech especially, there is cross-linguistic evidence that words at the edges of utterances are "acoustically more salient" than those in the middle (see [92]). Using an HPP procedure very similar to JA's, Seidl and Johnson [92] found that eight-month-old English-learning infants could segment words from utterance-initial and utterance-final positions, but not utterance-medial.

At the level of phonological phrases, Christophe et. al. showed that French-learning and Spanish-learning newborns can discriminate between lists of bisyllables excised from speech which (1) do (2) do not traverse phonological phrases [14, 15]. More significantly for segmentation of individual words, Gout, Christophe, and Morgan [38] habituated English-learning infants to bisyllabic words such as *paper*, then tested (following JA) whether they discriminated between sentences where the two syllables occurred within (*The scandalous paper...*) versus across (*The outstanding pay persuades...*) phonological phrases. They found that 10-month-olds did not and 13-month-olds did, suggesting that phonological phrase boundaries can only potentially be used for segmentation from late in the first year.

Coarticulation
In contrast to bracketing information from lengthening at prosodic edges, a clustering cue potentially availaable to infants is coarticulation. There is some evidence (see [52]) that coarticulation between two phonemes decreases when they are separated by a word boundary, that coarticulation cues help (adult) listeners place syntactic boundaries, and more generally that coarticulation can act as a "clustering" cue for adjacent elements. Johnson and Jusczyk [52] indirectly showed that coarticulatory information can be used as a segmentation cue by 8-month-old English-learning infants. Using a variant of SAN's stimuli where coarticulatory cues to word boundaries in the stream of syllables conflicted with TP cues, Johnson and Jusczyk found that subjects used coarticulatory over statistical cues.

[12]Writing "WS" or discussing English and French stress in similar terms is questionable, since French is is a quantity-insensitive, syllable-timed language where stress is less important than other prosodic factors. We stick with this notation for purposes of comparison with English.

Additional evidence comes from a study of 7.5-month-old English-learning infants by Curtin, Mintz, and Bird [20], who used a variant of SAN's stimuli to construct syllable streams (1) spoken normally (and hence coarticulated) (2) spliced together from the normal stream to be "miscoarticulated". Two groups of subjects were exposed to these streams, and while subjects in the appropriately coarticulated condition discriminated between familiar and novel syllable sequences, subjects in the miscoarticulated condition did not.[13] These studies show that in fairly different experimental conditions (JA and SAN), 8-month-olds can use coarticulatory cues for segmentation.

Like prosodic and statistical cues, then, acoustic cues provide imperfect evidence for word boundaries and are available to infants beginning at roughly 8 months.

3.3.3 Segmental distribution cues

"Statistical learning" refers to distributional information at the syllable level in the segmentation literature, but distributional information at the segmental level is potentially available to infants as well in phonotactic and allophonic patterns. For adults, there is evidence that phonotactics influence segmentation both at categorical (the "possible word constraint" [79]) and probabilistic levels [71, 104] by indicating where syllable boundaries are likely to be placed. There is also evidence [16] that at least in English, allophonic information often correlates with word boundaries (e.g. *night rate* vs. *nitrate*). When does sensitivity to segmental distribution develop in infants, and can they use distributional cues to segment?

Phonotactics

Several studies indicate that infants develop sensitivity to the phonotactics of their ambient language in the second half of their first year. Jusczyk and Luce [61] found that English-learning 9-month-olds listened longer to words from a list containing high-frequency phonotactic patterns than to those from a low-frequency list, but that 6-month-olds did not discriminate between lists. Importantly, this means 9-month-olds have some sensitivity to the relative frequency of *legal* phonotactic patterns which occur in their input, termed "probabilistic phonotactics". Nine-month-olds can also discriminate between words containing legal and illegal phonotactic patterns in their native language (specifically in English vs. Dutch input [55]), but such "categorical phonotactics" provide less useful information for word segmentation.

Mattys and Jusczyk [69] showed that 9-month-olds can use probabilistic phonotactics in segmentation tasks. Infants were exposed to CVC stimuli (*gaffe*) read in passages such that the stimuli onset and coda were part of consonant clusters (..C#CVC#C..). Passages differed only in whether the clusters provided good (.. *bean gaffe hold*..) or bad (.. *fang gaffe tine*..) phonotactic cues to segment the target word correctly. Subjects were then exposed to the CVC stimuli in isolation and tested using HPP; they looked significantly longer for stimuli from "good" than from "bad" conditions, and looked no longer for "bad" condition words than for control words (not included in passages).

Allophones

Less well-studied is whether infants can use allophonic cues for segmentation. Jusczyk, Hohne, and Bauman [58] used a variant of JA's procedure to familiarize English-learning infants with allophonic bisyllables (such as *nitrate*/*night rate*), then tested whether they discriminated between passages differing only by which allophonic variant was embedded. Nine-month-olds could not while 10.5-month-olds could, indicating that the ability to use allophonic cues in isolation develops between 9 and 10.5 months.

Taken together, these phonotactic and allophonic studies suggest that infants can segmental-level statistics for segmentation by 9-10.5 months, in addition to the syllabic-level statistics shown by SAN. This is surprising given that the dominant perceptual unit is the *syllable* until late in the first year (see 2.2), meaning infants may begin using information about the *distribution* of segments before representing words segmentally.

[13] Interestingly, adults in both conditions discriminated between familiar and novel sequences.

3.3.4 Perturbations, limitations of non-statistical cues

Under the hypothesis that infants segment by integrating non-statistical cues, there is a persistent chicken-and-egg problem: how would an infant *begin* this process? Since adults largely segment by prosodic cues and infants specialize to a native language first in the prosodic domain, the idea of "prosodic bootstrapping" seems plausible: infants use their language's dominant prosodic pattern to make a rough pass at segmentation, then build from there by integrating other cues.

This idea of "prosodic bootstrapping" is conceptually appealing, but rests on the assumption that the default stress pattern in a language will be most common in child-directed speech. In fact, Swingley ([95]) found that in a corpus of English child-directed speech, WS bisyllabic utterances were *more* common than SW utterances (token frequency),[14] meaning a learner would not have evidence for a trochaic bias. This is problematic not just because English-learning infants show a trochaic bias, but because they show it early and robustly in acquisition, suggesting that their learning mechanism can quickly derive it from limited input. Swingley found that when bisyllabic mutual information *and* S/W information were taken into account, a trochaic bias emerged cleanly and was robust when the amount of input was shrunk to one week's worth. These results also held for child-directed speech in Dutch, another strongly trochaic language. Since mutual information is equivalent to TPs, Swingley's study is a convincing demonstration that when faced with actual child-directed speech instead of the restricted set necessary to carry out laboratory experiments, a combination of statistical learning and prosodic cues rapidly succeeds where either one alone fails.

Swingley's study was only published recently (2005), but highlights a problem which is clear in hindsight: non-statistical learners are not in general robust. Such a hypothesized learner must quickly converge on the same prosodic pattern given very different inputs, some of which will violate any proposed language-specific cue. The learner must also know what to look for in the speech stream despite being pre-linguistic, but how? The standard answer is that the infant has innate knowledge specifying what to look for in the speech stream; SAN and later statistical work made a very different proposal: early segmentation is driven by domain-general pattern recognition of clustering information, which is (it was claimed) fast and largely independent of the input.

Support for a learner capable of rapidly deriving statistics over the input also comes from infants' ability to quickly switch hypotheses. In a recent elegant study, Thiessen and Saffran [97] showed that English-learning infants can quickly learn to switch from a trochaic bias to an iambic one. Two groups were habituated to lists of trochaic or iambic bisyllables, then listened to speech streams of concatenated trochaic or iambic nonsense words (as in SAN, but with stress), then were tested on words vs. part-words. The results indicated that both groups discriminated between words and part-words identically, meaning iambic subjects segmented using an iambic strategy and trochaic subjects using a trochaic strategy. This is striking evidence that English-learning infants must not (at least in early segmentation) rely too much on their trochaic bias, since it can be reversed in a few minutes! Thiessen and Saffran also tested whether 6-7 month-old infants, who had previously been shown to weigh statistical cues over (trochaic) stress cues when segmenting from SAN-style speech streams (see below: [96]), would segment differently after habituation to the iambic list. Subjects now segmented using an iambic strategy and weighed stress cues over statistical cues, meaning that brief habituation to iambic input both reversed infants' "trochaic bias" *and* changed their segmentation strategy.

Such results argue convincingly for a statistical component to infant word segmentation in addition to statistical cues; the question then becomes what their relative roles are. Experimentalists seem to basically agree on this point, and much work over the past decade has examined the relative role of *all* segmentation cues used by infants.

3.4 Integrating cues, generalizing across input

Work on infant segmentation has increasingly addressed the relative role of different proposed cues and how they are integrated over time, especially along the statistical/non-statistical axis.

[14] Owing to phrases like "who's this?" and "oh dear!"

3.4.1 Integrating non-statistical cues

Among non-statistical cues, the intuitive hypothesis is that prosodic cues outweigh non-prosodic cues in infant segmentation, as expected if prosody is the first domain in which infants have language-specific knowledge. In a first study checking this hypothesis, Mattys et. al. [70] tested the relative role of stress and phonotactics in segmentation by 9-month-old English-learning infants. Groups of infants were habituated to lists of CVC.CVC words stressed either initially or finally and with medial consonant clusters which usually occur either word-internally or between words in English. Using an HPP procedure, Mattys et. al. then tested whether infants segmented stimuli as whole words or not. The experiments used are somewhat involved, but indicate that while subjects were sensitive to phonotactic cues, but relied more on stress cues when both were present.

3.4.2 Weighing non-statistical and statistical cues

Several studies have examined the relative role of statistical and non-statistical cues in segmentation by infants, in particular stress vs. TPs. Johnson and Jusczyk [52] replicated SAN's methodology using 8-month-old infants habituated on two kinds of speech streams. Streams were identical to those used by SAN (concatenated trisyllabic nonsense words with no pauses between words), except that some streams now contained conflicting stress and TP cues for segmentation[15] and some contained conflicting coarticulation and TP cues. Both stress and coarticulatiory cues outweighted TP cues, leading Johnson and Jusczyk to tentatively conclude that speech cues outweigh statistical cues. Thiessen and Saffran [96] replicated this finding in 9-month-olds for stress vs. statistics, but then showed that its reverse holds for 7-month-old infants, who weighed TP cues over stress. Based on these findings, Thiessen and Saffran suggested that infants initially rely primarily on statistical cues for segmentation, then give increasing weight to non-statistical cues from 7 months on. The same authors recently showed [97] (discussed in 3.3.4) that 7-month-old English-learning infants switch to weighing stress above TPs after listening to a list of regularly-stressed words, suggesting that the weighting of cues at this age is malleable.

A study by Toro-Soto, Rodríguez-Fornells, and Sebastián-Gallés [100] using Spanish-speaking adults serves as a reminder that much more cross-linguistic study is needed before conclusions are reached on how segmentation works. They replicated Thiessen and Saffran's stimulus streams, except that different streams had word-initial, word-medial, word-final, random, or flat stress. Since stress in Spanish is largely penultimate, it was expected that word-medial stress would aid segmentation; instead word-medial subjects performed the worst, picking correct words significantly *below* chance. By contrast, initial-, final-, and flat-stress groups performed above chance and identically, suggesting that subjets were segmenting based on TPs alone, and word-edge stress information did not help. It is unclear how to account for these results, particularly for the word-medial group, where the type of native-langauge stress cues which help English speakers segment hindered Spanish speakers.

3.5 Generalizations

Ultimately infants must segment from variable input, often from multiple dialects or languages, and learn to generalize across variability in how a given word is produced. Some studies have begun to untangle these issues.

3.5.1 From sequences to words

Most studies discussed so far assume that the fact infants can recognize sequences of syllables in a new context means they are performing something like word segmentation, i.e. grouping syllables together as a linguistic unit. Is this true, and over what age does it develop? Evidence that the output of SAN-type experiments is indeed linguistic units was reviewed above (3.2.1). In an earlier study, Morgan and Saffran

[15] For example *pabiKUtibuDOgolaTUdatoPItibuDO...* which would be segmented into *pabiku*, *tibudo..* based on TPs and *kutibu*, *dogola...* based on stress.

[74] exposed 6-month-olds and 9-month-olds to series of syllables differing by whether they contained serial order information (about subsets of syllables) and by stress pattern (trochaic vs. iambic). The procedure is complicated, but eventually Morgan and Saffran concluded that "whereas 9-month-olds appear to be capable of integrating sequential and suprasegmental information in forming word-like (multisyllabic) phonological percepts, 6-month-olds are not", suggesting that the ability to cluster syllables together as words develops in this time range. Mattys and Jusczyk [68] checked whether JA's findings held using phonemes occurring across word boundaries ("cold ice") instead of the equivalent words ("dice"). They did not, and Mattys and Jusczyk eventually concluded that infants in JA-type studies are segmenting words rather than "recurring contiguous patterns".

3.5.2 Cross-linguistic carryover?

Given that infants' input can be multilingual, it is natural to ask to what extent infants' segmentation mechanisms generalize across languages. Although experiments discussed above (such as SAN) using artificial languages offer one perspective, because the languages used are simple and homogeneous it is not clear to what extent a *positive* result generalizes to natural languages.

At one extreme, experiments where English-learning infants were exposed to up to several hours of Chinese input (see 3.2.2) did not find evidence for new segmentation abilities. However, other studies have found that training on relevant properties of the new language quickly activates segmentation abilities not present in the native language: Höhle et. al. [44] found that French-learning 6-month-olds do not initially distinguish between non-word SW and WS stimuli produced by a German speaker, but do do so after 1 minute of training on SW or WS sequences alone.[16] Thiessen and Saffran ([97], see 3.3.4) found similar plasticity in 6 to 7-month-old English-learning infants, who switched to preferring iambic (instead of trochaic) bisyllables after brief exposure to a list of iambs. It is not clear whether the divergent results of the English/French/German studies and the English/Chinese studies is due to greater typological similarity between the European languages or that infants were not trained on specific properties of Chinese relevant for segmentation.

At the other, there is some evidence that segmentation abilities generalize across similar languages or dialects. Polka and Sundara [83] found (using JA's methodology) that 8-month-olds learning Canadian French showed evidence of segmentation when either Canadian French or European French stimuli were used, despite non-trivial phonological and prosodic differences between the dialects. Houston et. al. [47] (also using JA's methodology) found that English-learning and Dutch-learning infants could segment SW Dutch words from Dutch passages to the same degree. Since Dutch and English are similar prosodically but otherwise quite different, Houston et. al. hypothesized that infant segmentation abilities extend to languages with similar rhythmic structures. For both these findings, the key question is what "similar" means: without more cross-linguistic study, the hypothesis that segmentation abilities generalize more across languages/dialects that are more closely genetically-related cannot be ruled out.

3.5.3 Changing the target of segmentation

Most studies discussed above test whether infants can segment trisyllabic nonsense words (SAN) or nouns of 1-2 syllables (JA), usually made up of CV or CVC syllables, from some context, Natural language consists of words with different lengths, syllable contents, and parts of speech, and a few studies have checked how similar segmentation abilities are when different stimuli are used.

Word length

Using JA's methodology, Houston, Santelmann, and Jusczyk [48] examined whether 7.5-month-old English-learning infants could segment trisyllabic English nouns from fluent speech; they found that subjects seg-

[16] By contrast with non-native segmental distinctions, which are lost over infancy, French-learning children do not lose the ability to use stress in segmentation: Tyler [102] showed that French-speaking adults can use stress to segment an artificial language at the same level as Dutch-speakers.

mented ŚWȘ stimuli as words [17], but missegmented ȘWŚ stimuli, treating the last syllable as a separate word. Thus English-learning infants can segment trisyllabic nouns from fluent speech as early as bisyllabic nouns as long as they fit English's dominant trochaic rhythm.

Part of speech
Intuitively it seems that the part of speech of words used in segmentation experiments around 6-12 months should not matter, since infants have no syntactic abilities at this age. Höhle and Weissenborn used JA's methodology to check if German-learning infants could segment German closed-class elements (≈ function words) from fluent speech. Six-month-olds could not, but 7-9 month-olds could, which is about the same timescale as for nouns (in German, English, and Dutch). This indirectly argues for the importance of TPs in early segmentation, since function words are rarely spoken in solation.

However, Nazzi et. al. [76] found a striking developmental difference between nouns and verbs. They used JA's methodology, except that the SW and WS (English) words to be extracted from fluent speech were now verbs (rather than nouns). 13.5-month-old infants could segment bisyllables, but 10.5-month-olds could not, while JA found that 7.5-month-olds could segment SW nouns, indicating that (English) verb segmentation trails noun segmentation by at least 3 months. Nazzi et. al. suggest that this delay could be due to the different prosodic environments of verbs and nouns in spoken English, but regardless of the cause it highlights the fragility of early segmentation abilities.

Onset type
Intuitively, if most words begin with consonants, segmenting C-initial words should be easier than segmenting V-initial words. Several studies support this idea. Nazzi et. al. (above) found that V-initial WS verb stimuli were only segmented by 16.5 months, whereas WS C-initial verb stimuli were segmented by 13.5 months. For English nouns, Mattys and Jusczyk found that even monosyllabic V-initial nouns were not segmented until 13-16 months, compared to 6-7.5 months for C-initial nouns, a huge delay in language development terms. Thus, early segmentation skills seem to depend heavily on syllable type. However, a recent study by Seidl and Johnson (using the same methodology) [93] shows that 11-month-olds can segment V-initial monosyllables *if* they are in sentence-initial or sentence-final position; this again highlights that infants segment using a set of cues which are fragile individually but robust when combined.

The more general point that consonants and vowels are treated differently in segmentation has also been made in several studies by Mehler and colleagues [6, 73, 98] mostly using adult subjects (and so not reviewed here).

3.5.4 Words and forms in isolation

Putting aside segmentation from fluent speech, there is a simple null hypothesis as to how infants begin to segment: they hear highly frequent small chunks in isolation, then pick these chunks out of sentences. That highly-frequent chunks are often bigger than words ("come here") is not problematic; many of infants' first utterances are of this form. Is segmentation based on isolated chunks plausible given children's input? For English, Brent and Siskind [10] found that 9% of utterances to children in English child-directed speech are single words, and the order in which these words are acquired roughly correlates with their frequency. Researchers differ on whether this shows that isolated words are too infrequent to bootstrap from or that infants do make use of frequent words spoken in isolation[18], but it is clear that there is a core chicken-and-egg problem with *pure* bootstrapping from isolated words: the child needs to know some words to distinguish multi-word utterances from single words, and vice versa. The problem is much clearer for agglutinative languages or those with rich inflectional morphology.

However, words (or small chunks) in isolation are important for a different reason: for the experiments discussed so far to bear on lexical development, infants must *retain* the words they segment and be able to recognize them in novel contexts. Several studies suggest this ability develops during the relevant time range

[17] Ś=primary stress, Ș=secondary stress

[18] "[On Brent and Siskind's findings] Clearly, isolated words are abundant in the learning data and children do make use of them." [34]

(6-12 months). Houston and Jusczyk [46] familiarized 7.5-month-olds with pairs of words, then checked whether they could segment them from fluent speech 1 day later (using JA's methodology). They found that subjects did so only if the same speaker read the fluent speech and original words, implying that 7.5-month-olds' representations are not speaker-independent, and hence somewhat fragile. Furthermore, Houston, Tincoff and Jusczyk [49] found that when the same stimuli were used, 7.5-month-olds failed to segment after a one-week delay.

However, the duration and detail of word representations rapidly develops:[19] Houston and Jusczyk [45] found (using similar methodology) that 10.5-month-old English learning infants could generalize across speaker gender to segment, while 7.5-month-olds could not. Jusczyk and Hohne [57] found that English-learning 10-month-olds who were read children's stories could discriminate two weeks later between lists of words which occurred frequently or did not occur in stories. (Lists and stories were read by the same speaker.) Swingley [94] found that 11-month-old Dutch-learning infants listened longer to familiar words than non-words, but not if slight mispronunciations were made in familiar words.

The overall picture is that infants' segmentation abilities develop from fragile and rudimentary to robust over exactly the same timeframe ($\approx 7.5 - 11$ months) that their stored representations progress from fragile and perishable to robust and durable. One wonders how these processes are linked.

4 Conclusion

From work done to date, a tentative idea of how infants learn to segment can be suggested. Around 6-8 months, they begin to use distributional information at the syllabic level and language-specific prosodic cues. The representations they store are short-lived and are very detailed, so that infants cannot generalize across speakers or contexts. Over the next few months infants become able to integrate phonotactic and allophonic cues for segmentation in order of their frequency of occurrence. Words are first segmented later depending on their part of speech and syllable structure, or earlier if they are located at prosodic boundaries. The relative importance of cues varies as infants get older; there is some evidence that distributional cues (SL) are weighted higher than non-statistical cues in early segmentation(6-7 months), but lower by 9 months and thereafter. The overall picture is one of a host of cues, most of which depend on language-specific information but also reflect types of information (prosody, TPs, phonotactics..) common to all languages. Every cue (or even class of cues) is unreliable in isolation because of exceptions and significant variability in the input; it is only by integrating cues that infants can segment in increasing detail. Also important is the *fragility* of most mechanisms infant use to segment when used in isolation. Until a few months into learning to segment, making minor changes to input stimuli erase the findings of most studies discussed.

In all, this picture is both extremely interesting and extremely tentative. As the few studies done on languages besides English (and to a lesser extent Dutch and French) suggest, more cross-linguistic study is badly needed to form a robust picture of how infants segment. It striking that researchers at one point hypothesized a general, cross-linguistic "trochaic bias" in infant speech perception, which in hindsight was simply a result of most research being on infants learning West Germanic languages. More research is also very much needed to examine to what extent results using artificial languages and/or with adult subjects generalize to infant perception of natural language. Adults and artificial languages are obviously much easier to study than infants and natural languages, but without direct comparative study it is unclear how much the substantial literature studying the former can be generalized to the latter. "Trochaic bias" findings are again instructive: whereas English-speaking adults primarily segment using a trochaic MSS, this turned out not to be the case for English-learning infants.

While the statistical vs. non-statistical debate has been productive in stimulating research from new perspectives and on how cues are integrated, it has perhaps distracted from fundamental questions having to do with the specific problem of how infants segment. By focusing on mechanisms which are hypothesized to be the same between adults and children, cross-linguistically, and domain-general, SL research sometimes glosses over the important questions discussed above. This is true for example in artificial grammar studies,

[19]There is in fact a significant literature and much debate on the development of infants word representations; the studies cited are just a relevant selection for current purposes.

which are fascinating if one takes their results to be indicative of how natural languages are learned, and easy to dismiss if not, since very little work has been done examining how much they generalize. On the non-SL side, research tends to be done incrementally in tightly controlled situations using a very small set of natural languages, making it equally difficult to generalize (but for the opposite reason). It is not surprising that some of the most interesting studies discussed here integrate *both* approaches.

In all, infant word segmentation remains a rich and fascinating topic about which some basics are known, but much more remains to be discovered.

References

[1] 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, 1998.

[2] R.N. Aslin, J.R. Saffran, and E.L. Newport. Statistical learning in linguistic and nonlinguistic domains. In *The emergence of language*, pages 359–380. Lawrence Erlbaum, 1999.

[3] R.N. Aslin, J.Z. Woodward, N.P. LaMendola, and T.G. Bever. Models of word segmentation in fluent maternal speech to infants. In J.L. Morgan and K. Demuth, editors, *Signal to syntax: bootstrapping from speech to grammar in early acquisition*, pages 117–134. Lawrence Erlbaum, 1996.

[4] E.O. Batchelder. Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83(2):167–206, 2002.

[5] J. Bertoncini, R. Bijeljac-Babic, P.W. Jusczyk, L.J. Kennedy, and J. Mehler. An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, 117(1):21–33, 1988.

[6] L.L. Bonatti, M. Peña, M. Nespor, and J. Mehler. Linguistic constraints on statistical computations. *Psychological Science*, 16(6):451–459, 2005.

[7] M.R. Brent. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning Journal*, 34(1):71–105, 1999.

[8] M.R. Brent. Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences*, 3(8):294–301, 1999.

[9] M.R. Brent and T.A. Cartwright. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, pages 93–125, 1996.

[10] M.R. Brent and J.M. Siskind. The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):33–44, 2001.

[11] P. Cairns, R. Shillcock, N. Chater, and J. Levy. Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33(2):111–153, 1997.

[12] G. Casillas. The insufficiency of three types of learning to explain language acquisition. *Lingua*, 118(4):636–641, 2008.

[13] M.H. Christiansen, J. Allen, and S. Seidenberg. Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13(2):221–268, 1998.

[14] A. Christophe, E. Dupoux, J. Bertoncini, and J. Mehler. Do infants perceive word boundaries? an empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, 95:1570–1580, 1994.

[15] A. Christophe, J. Mehler, and N. Sebastián-Gallés. Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2(3):385–394, 2001.

[16] K.W. Church. *Phonological parsing in speech recognition.* Kluwer, Norwell, MA, 1987.

[17] R.A. Cole and J. Jakimik. Understanding speech: How words are heard. In G. Underwood, editor, *Strategies of information processing*, pages 66–116. Academic Press, New York, 1978.

[18] R.A. Cole and J. Jakimik. How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 67:965–970, 1980.

[19] R.A. Cole and J. Jakimik. A model of speech perception. In R.A. Cole, editor, *Perception and production of fluent speech*, pages 133–163. Kluwer, Dordrecht, 1980.

[20] S. Curtin, T.H. Mintz, and D. Byrd. Coarticulatory cues enhance infants recognition of syllable sequences in speech. In *Proceedings of the 25th Annual Boston University Conference on Language Development*, pages 190–201, 2001.

[21] A. Cutler. Exploiting prosodic probabilities in speech segmentation. In G.T.M. Altmann, editor, *Cognitive Models of Speech Processing*, pages 104–121. MIT Press, Cambridge, MA, 1991.

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

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

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

[25] A. Cutler, J. Mehler, Norris D.G., and J. Segui. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4):385–400, 1986.

[26] A. Cutler, J. Mehler, D. Norris, and J. Segui. A language-specific comprehension strategy. *Nature*, 304:159–160, 1983.

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

[28] C. de Marcken. Linguistic structure as composition and perturbation. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 335–341, 1996.

[29] 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(2):202–225, 1997.

[30] P.D. Eimas. Infant speech perception: processing characteristics, representational units, and the learning of words. In *The psychology of learning and motivation*, volume 36, pages 127–169. Academic Press, 1997.

[31] P.D. Eimas. Segmental and syllabic representations in the perception of speech by young infants. *Journal of the Acoustical Society of America*, 105:1901–1911, 1999.

[32] J.L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[33] J.A. Fodor, M.F. Garrett, and S.L. Brill. Pi ka pu: The perception of speech sounds by prelinguistic infants. *Perception and Psychophysics*, 18(2):74–78, 1975.

[34] T. Gambell and C. Yang. Word segmentation: quick but not dirty. Ms, Yale University.

[35] R.L. Gómez. Statistical learning in infant language development. In G. Gaskell, editor, *Oxford Handbook of Psycholinguistics*. Oxford University Press, Oxford, 2007.

[36] R.L. Gómez and L.A. Gerken. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5):178–186, 2000.

[37] J.V. Goodsitt, J.L. Morgan, and P.K. Kuhl. Perceptual strategies in prelingual speech segmentation. *J Child Lang*, 20(2):229–52, 1993.

[38] A. Gout, A. Christophe, and J.L. Morgan. Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, 51(4):548–567, 2004.

[39] Z.S. Harris. From phoneme to morpheme. *Language*, 31(2):190–222, 1955.

[40] J. Hayes and H. Clark. Experiments in the segmentation of an artificial speech analog. In J.R. Hayes, editor, *Cognition and the Development of Language*, pages 221–234. Wiley, New York, 1970.

[41] R.A. Hayes, A. Slater, and E. Brown. Infants' ability to categorise on the basis of rhyme. *Cognitive Development*, 15(4):405–419, 2000.

[42] J. Hillenbrand. Perceptual organization of speech sounds by infants. *Journal of Speech, Language and Hearing Research*, 26(2):268–282, 1983.

[43] S.A. Hockema. Finding words in speech: An investigation of American English. *Language Learning and Development*, 2(2):119–146, 2006.

[44] B. Höhle, R. Bijeljac-Babic, T. Nazzi, B. Herold, and J. Weissenborn. The development of language-specific prosodic preferences during the first half-year of life and its relation to later lexical development: evidence from German and French. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007.

[45] D.M. Houston and P.W. Jusczyk. The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5):1570–1582, 2000.

[46] DM Houston and PW Jusczyk. Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Percepttion and Performance*, 29(6):1143–54, 2003.

[47] D.M. Houston, P.W. Jusczyk, C. Kuijpers, R. Coolen, and A. Cutler. Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin and Review*, 7(3):504–509, 2000.

[48] D.M. Houston, L.M. Santelmann, and P.W. Jusczyk. English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes*, 19(1):97–136, 2004.

[49] D.M. Houston, R. Tincoff, and P.W. Jusczyk. 7.5-month-olds' memory for words after a 1-week delay. In D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk, editors, *Jusczyk Lab Final Report*. 2003. http://hincapie.psych.purdue.edu/Jusczyk (accessed 3/08).

[50] M.A. Hunter and E.W. Ames. A multifactor model of infant preferences for novel and familiar stimuli. In *Advances in infancy research*, volume 5, pages 69–95. Ablex, Norwood, 1988.

[51] E. K. Johnson and P.W. Jusczyk. Exploring statistical learning by 8-month-olds: the role of complexity and variation. In D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk, editors, *Jusczyk Lab Final Report*. 2003. http://hincapie.psych.purdue.edu/Jusczyk (accessed 3/08).

[52] E.K. Johnson and P.W. Jusczyk. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4):548–567, 2001.

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

[54] P.W. Jusczyk and C. Derrah. Representation of speech sounds by young infants. *Developmental Psychology*, 23(5):648–54, 1987.

[55] 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(3):402–420, 1993.

[56] P.W. Jusczyk, M.B. Goodman, and A. Baumann. Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, 40(1):62–82, 1999.

[57] P.W. Jusczyk and E.A. Hohne. Infants' memory for spoken words. *Science*, 277(5334):1984, 1997.

[58] P.W. Jusczyk, EA Hohne, and A. Bauman. Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics*, 61(8):1465–1475, 1999.

[59] P.W. Jusczyk, D.M. Houston, and M. Newsome. The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3-4):159–207, 1999.

[60] P.W. Jusczyk, A.M. Jusczyk, L.J. Kennedy, T. Schomberg, and N. Koenig. Young infants' retention of information about bisyllabic utterances. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4):822–836, 1995.

[61] 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(5):630–645, 1994.

[62] P. Keating. Phonetic encoding of prosodic structure. In J. Harrington and M. Tabain, editors, *Speech production: models, phonetic processes, and techniques*, pages 167–186. Psychology Press, New York, 2006.

[63] D.G. Kemler Nelson, P.W. Jusczyk, D.R. Mandel, J. Myers, A. Turk, and L. Gerken. The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18(1):111–116, 1995.

[64] M. Korman. Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language*, 5:44–45, 1984.

[65] IY Liberman, D. Shankweiler, FW Fischer, and B. Carter. Explicit syllable and phoneme segmentation in young children. *Journal of Experimental Child Psychology*, 18:201–212, 1974.

[66] B. MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum, 2000.

[67] GF Marcus, S. Vijayan, S. Bandi Rao, and PM Vishton. Rule learning by seven-month-old infants. *Science*, 283(5398):77–80, 1999.

[68] S.L. Mattys and P.W. Jusczyk. Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27(3):644–655, 2001.

[69] S.L. Mattys and P.W. Jusczyk. Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2):91–121, 2001.

[70] 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(4):465–494, 1999.

[71] J.M. McQueen. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39(1):21–46, 1998.

[72] J. Mehler, E. Dupoux, and J. Segui. Constraining models of lexical access: the onset of word recognition. In *Cognitive models of speech processing*, pages 236–262. MIT Press, Cambridge, MA, 1990.

[73] J. Mehler, M. Peña, M. Nespor, and L. Bonatti. The "soul" of language does not use statistics: reflections on vowels and consonants. *Cortex*, 42:846–854, 2006.

[74] J.L. Morgan and J.R. Saffran. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66(4):911–936, 1995.

[75] 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(3):756–766, 1998.

[76] T. Nazzi, L.C. Dilley, A.M. Jusczyk, S. Shattuck-Hufnagel, and P.W. Jusczyk. English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, 48(3):279–298, 2005.

[77] T. Nazzi, G. Iakimova, J. Bertoncini, S. Frédonie, and C. Alcantara. Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54(3):283–299, 2006.

[78] R. Newman, J. Tsay, and P.W. Jusczyk. Segmentation of Chinese words after exposure to Chinese. In D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk, editors, *Jusczyk Lab Final Report*. 2003. http://hincapie.psych.purdue.edu/Jusczyk (accessed 3/08).

[79] D. Norris, J.M. Mcqueen, A. Cutler, and S. Butterfield. The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34(3):191–243, 1997.

[80] D. C. Olivier. *Stochastic grammars and language acquisition mechanisms*. PhD thesis, Harvard University, 1968.

[81] W.J. Ong. *The Presence of the Word: Some Prolegomena for Cultural and Religious History*. Yale University Press, 1967.

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

[83] L. Polka and M. Sundara. Word segmentation in monolingual and bilingual infant learners of English and French. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 1021–1024, 2003.

[84] F. Ramus, M. Nespor, and J. Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:265–292, 1999.

[85] A.N. Redlich. Redundancy reduction as a strategy for unsupervised learning. *Neural Computation*, 5(2):289–304, 1993.

[86] J.R. Saffran. Words in a sea of sounds: the output of infant statistical learning. *Cognition*, 81(2):149–169, 2001.

[87] J.R. Saffran. Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, 12(4):110–114, 2003.

[88] J.R. Saffran, R.N. Aslin, and E.L. Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, 1996.

[89] J.R. Saffran, E.L. Newport, and R.N. Aslin. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4):606–621, 1996.

[90] J.R. Saffran, E.L. Newport, R.N. Aslin, R.A. Tunick, and S. Barrueco. Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2):101–105, 1997.

[91] J.R. Saffran and E.D. Thiessen. Domain-general learning capacities. In E. Hoff and M. Shatz, editors, *Blackwell Handbook of Language Development*, pages 68–86. Blackwell, Oxford, 2007.

[92] A. Seidl and E.K. Johnson. Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, 9(6):565–573, 2006.

[93] A. Seidl and E.K. Johnson. Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, 35(01):1–24, 2008.

[94] D. Swingley. 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5):432–443, 2005.

[95] D. Swingley. Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50(1):86–132, 2005.

[96] E.D. Thiessen and J.R. Saffran. When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, 39(4):706–716, 2003.

[97] E.D. Thiessen and J.R. Saffran. Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3(1):73–100, 2007.

[98] J.M. Toro, M. Nespor, J. Mehler, and L.L. Bonatti. Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, 19(2):137–144, 2008.

[99] J.M. Toro, S. Sinnett, and S. Soto-Faraco. Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2):25–34, 2005.

[100] J.M. Toro-Soto, A. Rodríguez-Fornells, and N. Sebastián-Gallés. Stress placement and word segmentation by spanish speakers. *Psicológica*, 28:167–176, 2007.

[101] J Tsay and P.W. Jusczyk. Detection of words in fluent Chinese by English-acquiring and Chinese-acquiring infants. In D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk, editors, *Jusczyk Lab Final Report*. 2003. http://hincapie.psych.purdue.edu/Jusczyk (accessed 3/08).

[102] M.D. Tyler. French listeners can use stress to segment words in an artificial language. In *Proceedings of the 11th Australian International Conference on Speech Science and Technology*, pages 222–227, 2006.

[103] J. van de Weijer. *Language input for word discovery*. PhD thesis, Katholieke Universiteit Nijmegen, 1998.

[104] M.S. Vitevitch and P.A. Luce. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3):374–408, 1999.

[105] N. Warner and T. Arai. Japanese mora-timing: A review. *Phonetica*, 58:1–25, 2001.

[106] J.G. Wolff. The discovery of segments in natural language. *British Journal of Psychology*, 68:97–106, 1977.