Cognitive Psychology 67 (2013) 130–150

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Words and possible words in early language acquisition

Erika Marchetto [a,b], Luca L. Bonatti [c,*]

[a] SISSA/ISAS, via Bonomea 265, Trieste, Italy
[b] Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), Ecole Normale Supérieure, Paris, France
[c] ICREA, Universitat Pompeu Fabra, Barcelona, Spain

ARTICLE INFO

Article history:
Accepted 8 August 2013

Keywords:
Artificial language learning
Development of morphosyntax
Possible words
Lexical acquisition

ABSTRACT

In order to acquire language, infants must extract its building blocks—words—and master the rules governing their legal combinations from speech. These two problems are not independent, however: words also have internal structure. Thus, infants must extract two kinds of information from the same speech input. They must find the actual words of their language. Furthermore, they must identify its *possible* words, that is, the sequences of sounds that, being morphologically well formed, *could* be words. Here, we show that infants' sensitivity to possible words appears to be more primitive and fundamental than their ability to find actual words. We expose 12- and 18-month-old infants to an artificial language containing a conflict between statistically coherent and structurally coherent items. We show that 18-month-olds can extract possible words when the familiarization stream contains marks of segmentation, but cannot do so when the stream is continuous. Yet, they can find actual words from a continuous stream by computing statistical relationships among syllables. By contrast, 12-month-olds can find possible words when familiarized with a segmented stream, but seem unable to extract statistically coherent items from a continuous stream that contains minimal conflicts between statistical and structural information. These results suggest that sensitivity to word structure is in place earlier than the ability

* Corresponding author. Address: ICREA, Universitat Pompeu Fabra, C. Roc Boronat, 138, Edifici Tanger, 55.110, 08018 Barcelona, Spain.
E-mail address: lucabonatti@mac.com (L.L. Bonatti).

0010-0285/$ - see front matter
http://dx.doi.org/10.1016/j.cogpsych.2013.08.001

to analyze distributional information. The ability to compute nontrivial statistical relationships becomes fully effective relatively late in development, when infants have already acquired a considerable amount of linguistic knowledge. Thus, mechanisms for structure extraction that do not rely on extensive sampling of the input are likely to have a much larger role in language acquisition than general-purpose statistical abilities.

1. Introduction

1.1. Words and word structure in speech

To master a language, a learner must acquire both the extended network of words composing its lexicon and the complex web of structural relationships that make the lexicon productive, allowing the speaker to enjoy the infinite creativity unique to a human language (Chomsky, 1957). All children throughout the world accomplish these tasks in an incredibly short period of time, considering the complexity of a natural language. Their database is made up of samples of speech uttered under many different conditions, by many different speakers, with different intentions, purposes, and means of expression. Yet they converge to a final stable state within a few years. Thus, they must possess efficient means to extract the words of their language and their structural relationships from these speech snippets. What such means are, and what the relative importance of each of them is, are matters of intensive research.

Several layers of complexity make these tasks difficult. Speech contains no obvious marks of segmentation between words. So in order to find words, infants must break continuous speech into segments that will eventually become words in their languages. An important contribution towards the solution of this problem may come from infants' proven ability to extract some statistical relationships from speech. Notably, at 8 months of age, infants can spot words on the basis of their absolute frequency of occurrence (Jusczyk & Aslin, 1995), an ability that could help them mark some fixed points to break into continuous speech (Lew-Williams, Pelucchi, & Saffran, 2011). At around the same age or earlier (Thiessen & Erickson, 2013), they can compute the conditional frequencies with which syllables immediately follow each other (or their adjacent transitional probability; hereafter, TPs) to begin segmenting an artificially created continuous stream into word-like units (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). They can also exploit such distributional information when exposed to real speech (Pelucchi, Hay, & Saffran, 2009), suggesting that TP computations might be involved in real-world language acquisition (Lany & Saffran, 2010). In spite of these suggestive findings, we still do not know how strong infants' sensitivity to TPs is and whether it is strong enough to exploit real statistical differences in real languages, which are much less marked than those tested in the laboratory (Yang, 2004). Nor do we know whether sensitivity to such measures occurs at the right time in development. Undoubtedly, adults' finesse at detecting statistical relationships is exquisite (Vouloumanos, 2008). Timing is crucial in development, however: the acute sensitivity to small differences in statistical relationships needed to segment real speech is of little use for language acquisition if it comes when the problem that it should help solve—word segmentation—is already well on its way towards a solution.

Importantly, word segmentation is only a small step towards the construction of a lexicon. Words are not just bare syllable chunks. They are often composed of roots with added prefixes, as the word *prefix* exemplifies, suffixes, as *suffixes* exemplifies, or even simultaneous prefixes and suffixes, as *downhearted* or *embolden* exemplify. In short, words have internal structure, defined among adjacent and nonadjacent subparts. Their combinations define what *is* in the lexicon, but also what *could* be in the lexicon. To put the issue more abstractly, actual words are fundamental for the knowledge of a language, but so is the notion of a *possible word*—the general scaffolding of lexical items that contributes to language productivity (Pinker, 1991).[1]

Already at her very first steps into language, the learner faces the problem of finding both actual and possible words inextricably intermingled. However, to date, we know little about how and when infants form the notion of a possible word, nor whether they do it with the same kind of computational resources allegedly exploited to split the continuum into its composing segments. We know that at the very early stages of their production, children show a mastery of relatively abstract syntactic and morphosyntactic properties (Guasti, 2002; Lidz, Waxman, & Freedman, 2003), not only when they generate words that actually exist in their languages, but also when they make mistakes producing nonexistent, but legal words (e.g., Legate & Yang, 2007). We do not know, however, how such productive knowledge of possible words appears nor how it relates to the acquisition of actual words. We do know that very early in development infants can extract rules in different situations, but the role of this ability in the acquisition of linguistic competences in language acquisition is not well understood (Gerken, 2006; Gerken & Bollt, 2008; Gómez & Gerken, 1999; Johnson et al., 2009; Marchetto & Bonatti, 2013; Marcus, Vijayan, Rao, & Vishton, 1999).

1.2. Finding possible words in early language acquisition

Much research in language acquisition holds or implies the view that rules in language arise from an inductive process requiring a considerable amount of vocabulary and of variability within vocabulary (e.g., Altmann, 2002; Bates & Elman, 1996; Christiansen, Onnis, & Hockema, 2009; Gómez, 2002; Gómez & Maye, 2005; Onnis, Waterfall, & Edelman, 2008; Reali & Christiansen, 2005; Seidenberg, 1997). Infants first acquire many similar words, and from that database they extract information about their structure (Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2008). For reference, we will call this picture the *empiricist model* of language acquisition. According to this view, in principle, language acquisition is no different from other processes of statistical learning: learners only need data and some general-purpose mechanism of data analysis in order to acquire language (e.g., Altmann, 2002; Bates & Elman, 1996; Christiansen et al., 2009; Gómez, 2002; Gómez & Maye, 2005; Laakso & Calvo, 2011; Onnis et al., 2008; Reali & Christiansen, 2005; Seidenberg, 1997), perhaps aided by primitive perceptual biases (Aslin & Newport, 2012). As Bates and Elman wrote, the imperfections due to the impoverished nature of linguistic input, often considered the sign of the presence of rich linguistic structures in the learner, "wash out with a large enough sample" (Bates & Elman, 1996, p. 1849).

Despite the attractiveness and simplicity of this model, the induction of structure from a large database is not necessarily a winning strategy for an infant learner. One word is one word, but one rule is thousands of words. In learning a lexicon that in a few months will grow spectacularly, infants would benefit greatly from grasping possible words as early as they can, projecting word structure without waiting for the lexicon to grow beyond their memory and their cognitive limitations. Thus, instead of extensively exploring evidence about words to induce their structure, infants may benefit from being less conservative and project hypotheses about possible words after acquiring few real words, deploying a learning procedure less akin to a rational assessment of available evidence than to guessing from exemplars (Bonatti, 2008). Several studies involving adults and infants suggest that mechanisms not reducible to the tracking of statistical distributions are present in language acquisition. Thus, adults engaged in a word-learning task tend to deploy a one-trial, "fast-mapping" strategy to fixate the meaning of novel words (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013), a strategy that requires neither extensive encounters with word–object pairings nor statistical sampling of such encounters. Likewise,

[1] Norris, McQueen, Cutler, and Butterfield (1997) proposed the "possible word contraint", as a strategy to segment speech. According to it, the activation of a putative word candidate diminishes if its selection leaves neighboring segments that cannot be words (e.g., single consonants). Given a particular speech sequence, this constraint helps identifying which chunks of a really heard sequence may be selected as actual words. We use the concept in a different way. We consider the notion of *possible word* to refer to the disposition to accept novel sequences as belonging to the lexicon if they are legal combinations.

already at around 18 months, infants quickly learn novel words by means of elementary logical inferences that neither resemble nor require statistical computations (Halberda, 2003; Spiegel & Halberda, 2011). Even in highly idealized speech environments such as artificial language experiments, there is evidence that learners find words and word structure by means of different mechanisms that are sensitive to different features of the stimulus and are computing different functions. Peña, Bonatti, Nespor, and Mehler (2002) showed that adult participants can find words inside a continuous speech stream by computing statistical relationships among nonadjacent syllables, but are unable to extract information about the structures of these words when exposed to the same continuous familiarization, regardless of how long the familiarization is. Yet they are quickly capable of extracting the item structure once even a very short stream contains minimal, even subliminal, segmentation cues.[2] Rather than helping, more experience is detrimental to this process of structure extraction. Indeed, a longer exposure to speech streams, which would give the learner more opportunities to extract statistical relations, hinders the detection of generalizations (Endress & Bonatti, 2007). These facts are difficult to account for assuming that the acquisition of lexical knowledge is exclusively based on a single inductive process requiring extensive data gathering (Endress & Bonatti, 2013). Instead, adult learners seem to look for very specific information in order to either extract words or find the possible words of a language. For example, adults can find words inside an artificial stream of consonant–vowel syllables by computing TPs among the consonants, but they cannot find the structural relationship inside these words on the basis of the same computations. Conversely, they can find word structure by computing relationships over the vowels, but cannot use the same relationships to find the words (Bonatti, Peña, Nespor, & Mehler, 2005; Toro, Nespor, Mehler, & Bonatti, 2008). Remarkably, even young infants are sensitive to the same differences at the beginning of their word-learning journey (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Pons & Toro, 2010). All such considerations suggest that there is more to language acquisition than a cold analysis of the data by means of statistical computations. Words and possible words may be acquired by different learning strategies that respond to different aspects and different doses of the available experience. Recent adult neuroimaging evidence supports this conclusion, suggesting that the circuits involved in the projection of lexical rules and in the acquisition of lexical elements are different (De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Levi, 2007; De Diego-Balaguer, Fuentemilla, & Rodriguez-Fornells, 2011; Mueller, Bahlmann, & Friederici, 2008).

Here, we explore the hypothesis that at its onset language acquisition works differently from the process suggested by the empiricist model: that even before acquiring an extensive repertoire of words, infants try to construct the possible words of their language, so as to build a generative lexicon that extends beyond their experience with real words. If infants try to construct a notion of possible words while building their lexicon, then the question of the relative importance of actual and possible words arises. We propose that the most basic notion in lexical development is not the acquisition of novel words by means of statistical computations over a corpus of speech streams, but the projection of hypotheses about the possible words of a lexicon. We test this hypothesis by studying how 12- and 18 month-old infants—two crucial ages in lexical development—find words and possible words when exposed to artificial language streams. If our hypothesis is correct, then one would expect that infants develop a notion of possible words earlier than the moment at which they possess sufficiently refined

[2] The results by Peña et al. (2002) have often been criticized as an effect of phonological confoundings, rather than as proof of sensitivity to the structural properties of words in a segmented stream, because the consonants of the words in some of their AXC languages presented a plosive–liquide–plosive sequence (Gómez & Maye, 2005; Newport & Aslin, 2004; Perruchet, Tyler, Galland, & Peereman, 2004). However, such criticisms stem from an incomplete analysis of the original research, which already contained appropriate phonological controls (see Peña et al., 2002, fn. 17; see also Bonatti, Nespor, Peña, & Mehler, 2006). There is no denying that phonological factors affect word segmentation (Onnis, Monaghan, Richmond, & Chater, 2005), as it could not be otherwise. However, by themselves, they are not sufficient to explain adults' sensitivity to structure (see Endress & Bonatti, 2007, pp. 270 ff.). In any case, the material of the current experiments is not affected by such criticisms. The words have no common phonological pattern. Furthermore, we compare responses to the same test items by varying only the presence or absence of segmentation indices in the familiarization streams. Hence, any possible low-level phonological effect is factored into the design.

statistical tools to fruitfully extract words from speech. By comparing infants at these two ages, we can also explore this prediction.

1.3. Words and possible words in early development

Artificially synthesized speech is well suited to explore our question, because of the control it offers over many features of the stimuli, such as the possibility to eliminate prosody, to manipulate the speech rate, to create exact statistical relationships among syllables, or to introduce precise elements of segmentation. One way to probe sensitivity to real and possible words in early language acquisition is to create situations in which the same experimental stimuli contain diverging information. Suppose infants listen to artificial speech streams from which they could both identify some snippets on the basis of their statistical coherence and find word-internal structure that cannot be extracted by computing the same statistical measures. The way the conflict is solved will indicate the relative importance that possible words and real words have in early language acquisition. In the following experiments, we created such streams by merging and adapting to infants two paradigms introduced by Peña et al. (2002) and Aslin et al. (1998). Peña et al. (2002) exposed adults to streams of trisyllabic sequences characterized by internal nonadjacent transitional probabilities of 1 and a varying middle syllable. Following standard practice, we call such statistically coherent items *words*, although obviously this definition has only a limited relationship with real words in a natural language. Participants could use nonadjacent distributional information to identify such words. However, the words also came in families. For example, *PULIKI*, *PUBEKI*, and *PURAKI* were words as identified by nonadjacent distributional information, but also shared the common structure *PU_KI*. Thus, not only could participants extract the actual words that occurred in the stream, but they could also identify their morphosyntactic structure. We will call languages constructed with these items *AXC languages*. After exposure to a stream of an AXC language, participants were tested for their preference for *words* vs. *part-words* or for *rule-words* vs. *part-words*. Rule-words were sequences of three syllables obtained by replacing the middle syllable of words with another syllable. They never occurred in the stream but were structurally similar to words because they shared the same morphological construction. Part-words were sequences of three syllables that occurred in the stream across word boundaries, but had no common constructions. Thus, AXC languages contain both statistical information favoring certain groups of syllables (words or part-words, to different degrees) and examples of words that may induce the projection of a rule potentially applying to unheard, novel words. Because rule-words never occurred in the stream, sensitivity to rule-words signals that the learner identified the structural aspects of word construction. Instead, because either words or part-words always had non-null TPs and frequencies, the ability to extract them from a stream signals that learners identified them by computing some measure of statistical information.

To create a conflict between sensitivity to possible words and statistically extracted words, we created AXC languages by chaining a few words with a common structure (and a nonadjacent TP of 1 between their first and last syllables), but, adopting a method introduced by Aslin et al. (1998) used here with a different logic, we repeated some of them so that some of the transitions between words would also turn out to be statistically highly coherent. Specifically, our streams contained only four words, but two of them were twice as frequent, so that the stream also contained trisyllabic sequences of adjacent syllables, spanning word boundaries, that occurred with a frequency as high as that of words and with high adjacent and nonadjacent TPs (.67). Thus, with LIMUFE and BAGASO twice as frequent as *ligafe* and *bamuso* (where uppercase letters indicate highly frequent words, and italic type indicates words with low frequency), the streams could contain snippets such as:

. . .. . .LIMUFEBAGASO*ligafe*BAGASO*bamusoligafe*BAGASOLIMUFEBAGASO. . .

Given the arrangement of the words in the stream, some trisyllabic sequences spanning two infrequent words, such as *musoli*, occurred infrequently. However, some other sequences spanning two frequent words occurred frequently. For example, the sequence LIMUFEBAGASO occurred frequently and, as a consequence, so did the sequence FEBAGA, which is contained within it.[3] Indeed, such a sequence occurred even more frequently than a low-frequency word such as *bamuso*. We call such sequences *high frequency (HF) part-words*.

In order to convey the logic of our experiments, it is worth commenting on the role of statistical measures in defining the notion of a word in artificial languages and on its relationship with words in natural languages. In natural languages, words are rich entities endowed with many layers of representations, spanning from phonology to meaning. By contrast, in artificial languages what counts as a word depends entirely on the statistical relationships among syllables. It is a presupposition of research using artificial speech that, at some level, such statistical relationships do contribute to creating those groups of sounds that we call *words* in natural languages. However, given the enormous distance between natural language words and 'words' induced by exposing participants to a few minutes of an artificial language, the relationship between the latter ones and the former ones is at most indirect and reduces to the fact that they both enjoy similar statistical properties: they both have TPs and absolute frequencies higher than other syllable sequences in their respective languages. In short, artificial languages can show how certain computations may allow the learner to single out some snippets of speech stream as *potential candidates* for items that might become 'real words' under the appropriate conditions. Going back to our stimuli with these considerations in mind, it is important to remember that words in our AXC languages became potential candidates for segmentation by virtue of their statistical properties, but also HF part-words possess most of the same statistical properties. Therefore, HF part-words may be equally good candidates for word segmentation as words themselves. In some respects, they are even better candidates: infrequent words have higher nonadjacent TPs than HF part-words, but have lower adjacent TPs. Although ours is the first study comparing infants' ability to compute adjacent and nonadjacent TPs in the same experiments, there is universal agreement that, if anything, for infants (and most likely for adults) adjacent relations should be more salient than nonadjacent relations. Thus, besides containing few words having a common internal structure, our streams also contain some part-words that could be perceived as highly statistically coherent, on a par with or better than some words, and with a level of statistical coherence much higher than any real word in natural languages.

In our experiments, we first familiarized infants with streams that differed in some crucial properties that varied across experiments, but maintained the same statistical structure as explained above. Then, we tested their preferences for rule-words or HF part-words (Table 1). Crucially, in our experiments, structure is present within words, defined as constancy of their first and last elements, but the variation of the middle syllable that could induce infants to create an abstract representation of the words' structure is minimal. Thus, if infants grasp possible words, they cannot do so by extracting statistical measures of inner variability which have been suggested to be crucial for infants to acquire nonadjacent structural patterns (Gómez, 2002). By contrast, HF part-words are favored by most statistical measures that we know infants can compute, particularly adjacent TPs and absolute frequency. If the representations that infants can form from a speech stream exclusively depend on such computations, then infants should always favor those items that are more strongly marked by the statistical relationships present in the input. If, instead, infants do not only look for actual words, but also try to form an abstract template of word structure from a few examples, then they should also be able to create a representation of the words that are possible, given the input, and hence accept rule-words as legal, although they have 0 frequency and 0 adjacent TPs.

We began our investigation by testing the boldest hypothesis we could formulate: infants extract words and possible words pretty much just as adults do, with the same kinds of stimuli, in the same input conditions, and hence most likely by means of the same processes (Peña et al., 2002). Thus, we

[3] It may be observed that, just as FEBAGA may be considered an HF part-word, so could all the *n*-syllabic strings such as FEBA, GAFEBA, LIGAFEBA, or LIGAFEBAGA, or GAFEBAGASO, etc. Indeed, in general, every string that spans several words without being aligned to them—no matter how long—could count as a part-word. We follow standard practice in the literature, comparing trisyllabic items against trisyllabic items, although this choice is ultimately arbitrary and may have experimental consequences. Indeed, infants and adults may be impaired by strings of various lengths (Hoch, Tyler, & Tillmann, 2012; Johnson & Tyler, 2010). Our research may be seen as largely independent of these issues, however, because we compare infants' ability to extract possible words or real words according to the nature of the familiarization stream they are exposed to, all else being equal. Thus, our main measure of interest concerns the relative differences in looking times to the same items as the conditions of familiarization vary.

Table 1
Words used to compose the familiarization streams of the experiments and test items (Rule-words/HF part-words). All caps indicate frequent items, small caps indicate infrequent words, and italics indicate null frequency items. Boldface indicates the structural composition of the words. The pound symbol indicates the position of the pause occurring in the part-words of some experiments. Nonadjacent TPs in words and rule-words was 1. Adjacent TPs in words ranged between 0.67 and 0.5 in HF words and between 0.33 and 0.5 in LF part-words. Adjacent TPs in rule-words was 0. Adjacent TPs in HF part-words ranged between 0.67 and 0.5.

Familiarization	Test items	
Words	Rule-words	HF part-words
BAMU**SO**	***ba***_li_***so***	GASO(#)LI
BAGA**SO**	***ba***_fe_***so***	MUFE(#)BA
LIMU**FE**	***li***_ba_***fe***	SO(#)LIMU
LIGA**FE**	***li***_so_***fe***	FE(#)BAGA

tested the hypothesis that encounters with a few words instantiating a morphosyntactic rule embedded in a stream with segmentation indices separating words should foster sensitivity to possible words, favoring rule-words over HF part-words. By contrast, exposure to a continuous stream with the same statistical properties should favor the detection of the most statistically coherent items, favoring HF part-words over rule-words. We tested infants of two age classes: 18 months and 12 months. At 18 months, infants are already firmly engaged with language production. This implies that they have already mastered the perception and production of most sounds of their language and that their lexicon already includes information about word structure (Lidz et al., 2003). So, we expect them to be able to extract either words or possible words once exposed to speech streams with properties that may allow them to trigger either statistical computations over the streams or the projection of generalizations from word examples. In contrast, at 12 months, infants are only at their first stages of speech production. They are still in the process of zooming into their natural language and completing the fixation of crucial language-specific phonological properties (Werker & Tees, 1983, 2002)—a precondition for developing an extensive natural language lexicon. Indeed, word learning proper is generally assumed to begin around the first birthday (Bloom, 2000), when infants can also tell the difference between potential content words and language-specific functors (Shi, Werker, & Cutler, 2006). However, evidence that infants of that age can master morphosyntactic relations or can extract nonadjacent relations between or within words, is absent or severely limited. Yet, 12-month-olds are surely capable of computing adjacent TPs under varied conditions (Fiser & Aslin, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Saffran et al., 1996). Thus, if the empiricist picture of language acquisition is correct, 12-month-olds should be able to find statistically defined snippets of an AXC stream, but might not be able to form any notion of possible words. If, instead, a prominent pressure driving the acquisition of a lexicon is the urge to find its possible words quickly and early, then we should find sensitivity to word structure even at that age. Therefore, the way 12-month-olds negotiate the conflict between HF part-words and rule-words may allow us to gauge the relative importance of words and possible words at the onset of vocabulary acquisition.

We first test infants with the task that, in principle, is harder at both ages: the ability to form a notion of a possible word from a few word examples. Experiments 1, 2, and 3 probe 18-month-olds' ability to distinguish rule-words from HF part-words after exposure to a segmented stream while controlling for some important aspects of the material. Experiment 4 will test the same ability in 12-month-olds. Then, we will test whether infants can solve the conflict posed by a continuous stream containing items favored by most statistical measures (HF part-words) and items favored by a common structural description (words). Experiment 5 tests 18-month-olds, and Experiments 6 and 7 tests 12-month-olds. To anticipate our results, we will show that at both ages infants are able to find possible words, but only at the later age do they possess sufficient statistical sophistication to identify statistically coherent items in the streams. We believe that these findings may turn the way we conceive of the process of lexical acquisition upside-down, as they show that possible words have developmental priority over words in the construction of a novel lexicon. They pose a serious challenge to the empirical picture of language learning and call for alternative ways to conceive of language acquisition, which we explore in the General Discussion.

2. Experiment 1

2.1. Method

2.1.1. Participants

Sixteen 18-month-old, full-term infants from Italian-speaking families with a minimum APGAR of 8 and no hearing or vision problems were retained for analysis (12 girls; mean age: 18 mo, 20 d; age range: 18 mo, 3 d to 19 mo, 14 d). We used stringent criteria for including participants and trials in our dataset. Infants were considered to be fussy if during familiarization or test they gave signs of discomfort while listening to the stimuli, such as making more than sporadic vocal emissions or frequently turning their heads towards their caretakers. We also excluded infants whose caretaker actively interfered with the experiment by talking to the infants or exhorting them to look at the stimuli during the test phase. Offline, we also excluded infants who looked longer than 65 cumulative s in more than two test trials (i.e., one min plus a conservative amount of time for potential experimental error) on the assumption that these infants were sticky-fixating our attractor stimuli instead of responding to the test sounds. These criteria were applied to all the experiments. According to these criteria, 22 infants participated but were excluded from analysis (20 because of fussiness, 2 because they exceeded maximum looking time criteria). Because the experimenter was unaware of the conditions of the experiment, the infants excluded because of fussiness were excluded before she could know whether they had behaved in accord with or in opposition to the hypothesis tested. Indeed, most excluded infants fussed before ever getting to the test phase.

2.1.2. Stimuli

To familiarize infants with the AXC language, we created an artificial speech stream by pseudo-randomly concatenating four trisyllabic nonsense sequences syllables. In these sequences, the *words* of the language, the first syllable always predicts the third syllable, whereas the middle syllable varies (Table 1). Thus the language could be described as containing two AXC families characterized by minimal variability. All individual syllables had the same frequency, but the words did not. The stream was constructed while respecting two main constraints: (i) two words were twice as frequent; and (ii) frequent words could occur twice in a row during the stream but without immediate repetitions. Words were separated by 200-ms pauses. The familiarization stream was synthesized with increasing and decreasing amplitude in the first and last 5 s, respectively, to avoid providing direct cues to word onsets. The stream lasted 2 min, 52 s.

Test items were HF part-words and rule-words, constructed as explained above. Rule-words had a frequency of 0, adjacent TPs of 0, but a non-adjacent TP of 1. By contrast, by virtue of the constraints used to build the stream, part-words spanning the boundaries of frequent words (HF part-words) had the fairly high adjacent and non-adjacent TPs of 0.67. The auditory material was synthesized with Mbrola (Dutoit, Pagel, Bataille, & Vreken, 1996), using the FR-2 database, flat prosody, 116-ms phoneme length and 200-Hz pitch.

2.1.3. Procedure

We implemented a modified version of the head-turn preference procedure (Nelson, Jusczyk, Mandel, & Myers, 1995). Infants sat on their caretaker's laps, in a dimly lit, quiet room with three monitors positioned in front of them and at each side. The caretakers listened to masking music and were instructed not to interact with infants during the experiment. The experimenter, unaware of its conditions, controlled the experiment looking at the infant from a monitor placed in a separate room. During familiarization, a visual stimulus (a movie of a moving hand) attracted infants towards the center, while the speech stream played. In the test phase, two trials for each of the eight test stimuli were presented in pseudo-random order, with the constraints that the same item could not be repeated in the immediately succeeding trial and that a maximum of three items of the same type could not occur in a row. In each trial, the attractor appeared at the center monitor first. After 1.5 s of continuous fixation, it disappeared and reappeared on one of the side monitors. As infants oriented towards it, the test item started playing repeatedly from loudspeakers hidden behind the monitors, until infants looked away for 2 s consecutively or else looked for more than 65 s cumulatively. Test items were repeated after 500 ms pauses. An Apple G5 controlled by PsyScope X (http://psy.cns.sissa.it/) ran the experiment. Infants' looking behavior was recorded by a camera hidden behind the center monitor. The camera also allowed the experimenter to control the procedure online, so that the test trials began and ended contingent upon the infant's looking behavior. Looking time was successively coded off-line. Looking times shorter than 1 s, or 3 SD beyond the general mean computed for each test item type, were excluded from further analysis. Offline coding was validated in two ways. First, a random group of 10 infants was selected and simultaneously coded by the main experimenter and by a second coder not involved in this research. Agreement, computed on each individual looking episode, was 100%. Then, a second random group of 10 infants was separately coded by a different experienced coder. Agreement, computed on each individual trial (16 per participant) was 98%.

2.2. Results and discussion

The logic of the head-turn procedure is that if infants extract information from some kind of familiarization (or habituation), then they will react differently to stimuli consistent with and stimuli inconsistent with that information. Generally, this reaction translates into a differential looking time to the two classes of stimuli, although the direction of this difference may change (e.g., Houston-Price & Nakai, 2004). Generally, longer looking time at the inconsistent items is considered to be a signal of surprise. In the current experiments, we assume this logic and consider that information has been successfully extracted from the familiarization if we observe a looking-time difference between HF part-words and rule-words. Although we are not committed to any particular direction of the looking time difference to the test items, we will follow the standard practice of interpreting longer looking times at one class of the stimuli as a sign of surprise, hence as an indication that the information infants extracted from the stream allowed them to identify the defining properties of the contrasting class.

Fig. 1A presents the results of Experiment 1. A repeated measure ANOVA, with Test Item Type (HF part-words, Rule-words) as a within-participant factor and participants as a random factor, revealed

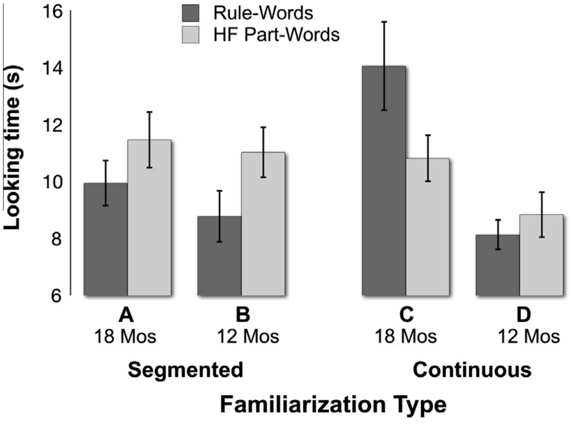

Fig. 1. Sensitivity to either structural or to statistical information as a function of familiarization type. Mean looking time (s) and SE for test items of 18-month-old and 12-month-old infants listening to either rule-words or High Frequency/High TP part-words, by familiarization type. Panels A (Experiment 1, 18-month-olds) and B (Experiment 4, 12-month-olds) report looking times after exposure to a segmented familiarization. Panels C (Experiment 5, 18-month-olds) and D (Experiment 6, 12-month-olds) report looking times after exposure to a continuous familiarization. Both 18- and 12-month-olds showed sensitivity to the possible words of a miniature language after exposure to a segmented stream. After exposure to a continuous stream, however, only 18-month-olds succeeded in identifying the most statistically coherent items of the language (HF part-words).

that infants looked longer while listening to HF part-words than to rule-words ($M_{\text{HF Part-words}}$ = 11.46 s, SEM = 1.0; $M_{\text{Rule-words}}$ = 9.94 s, SEM = 0.81, $F(1,15) = 5.37$, $P \leqslant 0.035$). Because rule-words never appeared during familiarization, while high-TP part-words did, this result suggests that infants extracted a generalization within words after exposure to a segmented stream containing few exemplars of such a generalization. By contrast, they disregarded sequences that occurred in the stream and that were favored by statistical measures, but that did not share the morphological construction of rule-words. This result closely patterns adults' behaviors after exposure to similar streams (Endress & Bonatti, 2007; Peña et al., 2002), suggesting that similar mechanisms may be involved. That infants at around that age are sensitive to morphological properties of words and phrases is known (Gertner, Fisher, & Eisengart, 2006; Gómez & Maye, 2005; Hirsh-Pasek & Golinkoff, 1996; Yuan & Fisher, 2009). What is remarkable is that, faced with a choice between two equally good candidates for words (one of which was a novel, but possible word, and another of which was a statistically highly favored candidate), infants clearly opted for the possible word.

One alternative explanation of these results could appeal to the difference between part-words listened during the familiarization stream and part-words presented during the test phase. The stream of Experiment 1 contained pauses between words. Therefore, every time infants listened to an HF part-word during familiarization, they also listened to a pause inside it. However, in the test phase part-words were uninterrupted trisyllabic groups. Thus, infants might have looked longer at part-words because they detected a mismatch between part-words during familiarization (with pause) and part-words during test (without pause). Experiment 2 tests this explanation.

3. Experiment 2

3.1. Method

3.1.1. Participants

Sixteen 18-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (7 girls; mean age: 18 mo, 26 d; age range: 18 mo, 7 d to 19 mo, 10 d). An additional 18 infants participated but were excluded from analysis (16 because of fussiness, and 2 because they exceeded maximum looking time criteria).

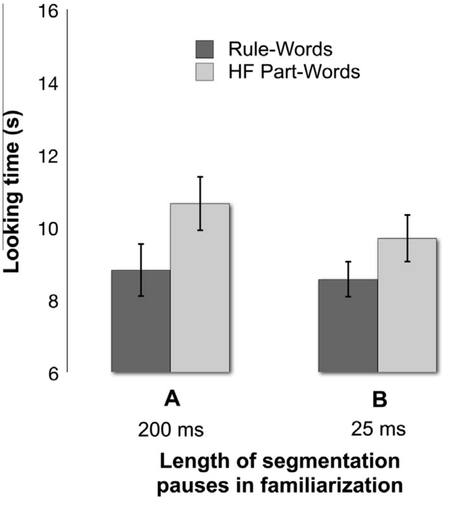

Fig. 2. The effect of segmentation marks inside the familiarization stream and in the test items. Mean looking time (s) and SE for test items of 18-month-old infants listening to either rule-words or High Frequency/High TP part-words. When both the familiarization and the HF part-word test items contained 200-ms pauses as segmentation marks (A: Experiment 2), infants identified the possible words of the miniature language. Likewise, when short, 25-ms pauses replaced the 200-ms pauses as segmentation marks in the familiarization (B: Experiment 3), 18-month-olds could still identify the possible words.

3.1.2. Stimuli and procedure

Infants were exposed to the same segmented familiarization stream as in Experiment 1. However, in the test phase the HF part-words were re-synthesized in order to include the segmentation pauses exactly as they appeared in the stream. Stimuli and procedure were otherwise identical to Experiment 1.

3.2. Results and discussion

Fig. 2A presents the results of Experiment 2. Infants looked longer toward HF part-words ($M_{\text{HF Part-words}}$ = 10.65s, SE = 0.76; $M_{\text{Rule-words}}$ = 8.81s, SE = 0.74, $F(1,15) = 5.46$, $P \leqslant 0.034$). This result confirms that infants' longer looking time towards HF part-words in Experiment 1 was not due to the discrepancies between the syllable sequences composing HF part-words heard during familiarization and those heard during test. Note also that assuming that pauses are treated as separate elements in computing TPs, the adjacent TPs between syllables and pauses inside HF part-words is 1. Therefore, if infants' behaviors were induced by adjacent TP computations over the transitions of elements in the stream, the statistical evidence should be even stronger in favor of the HF part-words including the pauses. Nevertheless, 18-month-olds preferred possible words with null frequency to items strongly favored by the statistical information contained in the stream. This result confirms the hypothesis that after being exposed to a segmented stream 18-month-olds naturally prefer legally defined, possible words to a series of syllables favored by many frequency measures.

In Experiments 1 and 2, pauses of 200 ms marked word boundaries, whereas in the studies by Peña et al. (2002) and Endress and Bonatti (2007) they were as short as 25 ms. We had decided to used large pauses to compensate for the differences in the acoustic properties of sounds presented via high-quality headphones (as used in experiments with adults) and those played in a room affected by unavoidable noises, such as those produced by the infants. While the precise nature of segmentation indices is not a focus of this study, however, it is important to establish whether infants can still grasp the morphosyntactic structure of words when the features of the input approach the connectedness of a real word stream. We now explore whether a clearly audible pause between words is a necessary element for infants to identify word-internal structure, or else, as happens with adults, minimal segmentation cues are sufficient to induce sensitivity to structure. Pilot studies showed that infants' reactions did not change with pauses of 100 ms and 50 ms. Therefore, in Experiment 3, we reduced the pauses between words to the bare minimum, replacing the 200-ms segmentation indices of Experiment 1 with 25-ms pauses.

4. Experiment 3

4.1. Method

4.1.1. Participants

Sixteen 18-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (7 girls; mean age: 18 mo, 7 d; age-range: 17 mo, 29 d to 19 mo, 9 d). An additional 18 infants participated but were excluded from analysis (16 because of fussiness, 2 because they exceeded maximum looking time criteria).

4.1.2. Stimuli and procedure

We synthesized a novel familiarization stream, built with exactly the same constraints as in Experiment 1 except for the fact that the 200-ms segmentation pauses were replaced by 25-ms pauses. Stimuli and procedure were otherwise identical to Experiment 1.

4.2. Results and discussion

Fig. 2B presents the results of Experiment 3. Infants looked longer towards HF part-words ($M_{\text{HF Part-words}}$ = 9.69s, SE = 0.67; $M_{\text{Rule-words}}$ = 8.55s, SE = 0.50, $F(1,15) = 6.75$, $P \leqslant 0.02$). This result shows that the exact nature of segmentation indices is not relevant for infants to project hypotheses about the possible words of an artificial language. Apparently, a varied set of segmentation indexes, even in connected speech and even if they are barely perceivable, triggers sensitivity to word structure. Again, this result closely patterns the behavior of adults tested by Peña et al. (2002) and Endress and Bonatti (2007).

Experiments 1–3 clearly show that at 18 months infants can create a representation of the possible words of a simple language, provided that they are exposed to a segmented stream including a few examples of its actual words. However, it is possible that the notion of possible word is not a fundamental primitive for language learning. It is reasonable to suppose, as suggested by the empiricist view of language acquisition, that infants first build a lexicon and then gradually form the notion of a possible lexicon, as a consequence of the accrued experience with their native language. If this acquisition sequence is correct, then younger infants exposed to a segmented stream containing a conflict between possible words and statistically favored syllable sequences may not be able to form a sense of its possible words as efficiently as 18-month-olds. To explore this possibility, Experiment 4 tests how 12-month-old infants, who are at a more primitive stage of natural language acquisition, can differentiate HF part-words from possible words after exposure to the same input conditions. Because we found no differences between Experiments 1 and 3, we used the materials of Experiment 1 to test 12-month-olds.

5. Experiment 4

5.1. Method

5.1.1. Participants

Sixteen 12-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (8 girls; mean age: 12 mo, 20 d; age range: 12 mo, 4 d to 13 mo, 0 d). An additional 22 infants participated but were excluded from analysis (18 because of fussiness, 4 because they exceeded maximum looking time criteria).

5.1.2. Stimuli and procedure

Stimuli and procedure were identical to Experiment 1.

5.2. Results and discussion

Fig. 1B presents the results of Experiment 4. Like their older peers, 12-month-olds looked longer while listening to HF part-words than to rule-words ($M_{\text{HF Part-words}} = 11.03$ s, SE = 0.91; $M_{\text{Rule-words}} = 8.79$ s, SE = 0.92, $F(1,15) = 5.34$, $P \leqslant 0.035$). The fact that both 18-month-olds and 12-month-olds behaved exactly like adults tested in similar (but obviously more complex) learning tasks (Endress & Bonatti, 2007; Peña et al., 2002) favors a view of continuity of the mechanisms of language acquisition across development. This result also shows that neither extensive experience with language (whether in life or during the experiment) nor extensive variability in the examples of a lexical rule are needed for infants to be able to project generalizations about possible words from a short speech stream. Such a result seems to be in contrast with the finding that infants exposed to little variability fail to learn nonadjacent dependencies, whereas they succeed in learning when extensively familiarized with variable material and with the relationships to learn (Gómez, 2002; Lany & Gomez, 2008). However, our results do not imply that infants, and especially older infants focused on the acquisition of the grammar of their languages, may not be helped in finding morphosyntactic patterns by extracting distributional information from the input; we know that, in certain cases, they are (e.g., van Heugten & Johnson, 2010). We demonstrated that this type of experience is not necessary for infants to form the notion of a possible word, which thus arises independently from the ability to take advantage of an extensive exposure to linguistic data. Indeed, in other cases variability may not be used at all in the process of acquisition of morphosyntactic patterns. For example, van Heugten and Johnson (2011) showed that Dutch infants learn some morphological dependencies in the inverse order with respect to that predicted by the variability hypothesis: infants learn diminutive dependency before the plural dependency, although corpus analysis shows that the plural dependency occurs in more variable contexts than the diminutive dependency. In short, our data, as well as more naturalistic results, show that the relationship between frequency, variability, and learning of patterns of generation of possible words is far more complex than the simple view that infants extract invariants from many variable examples would suggest (Johnson, 2011).

A possible alternative explanation to our results could invoke some uncontrolled baseline preference that infants might have for rule-words or part-words. To wit, they might show an inherent bias to look longer at HF part-words. Then, it may be imagined that exposed to a segmented stream, they do not modify their baseline preference for such items, thus generating the longer looking times for part-words than for rule-words observed in Experiments 1, 2, and 4 regardless of the structural properties of rule-words. A solid control for baseline preferences for words or part-words in experiments with infants exposed to artificial languages can be realized by modifying the statistical relations among syllables during familiarization in such a way that the status of the same tokens during the test phase switches according to the familiarization. Consider that tokens are either 'words' or 'part-words' only on the basis of the statistical relationships among the familiarization syllables. Thus, if with one familiarization some tokens are words and some others are part-words, with a novel different familiarization those words may become part-words in this context, and some part-words may become the new words. If infants prefer the same class of items (words or part-words) despite the fact that the very same tokens in one case are words and in the other case are part-words, then arguably possible baseline preferences cannot account for infants' looking behavior. The advantage of this technique, which several researchers in artificial statistical learning use (e.g., Aslin et al., 1998; Thiessen & Erickson, 2013), is that it exerts control over the specific syllable sequences used while allowing clear non-null predictions. Unfortunately, it cannot be used to control for baseline preferences between rule-words and HF part-words: there is no way to modify the familiarization so as to invert the role of HF part-words and rule-words. Simply put, part-words are essentially deprived of structure, and hence can never become rule-words. However, sequences of syllables that never appeared together in the stream (technically, *non-words*) could. Thus, a less perfect manipulation that still controls for possible baseline preferences consists of constructing familiarizations that maintain the token items in the test phase constant, as much as possible, but makes them become either rule-words or non-words by varying the syllable sequences in the streams. Marchetto and Bonatti (2013) did that with material almost identical to the current test items. They tested 12-month-olds by creating one segmented stream in which the rule-words were exactly those we used in the current experiments, whereas the non-words selected in the test phase were *sogali*, *femuba*, *bafemu*, and *lisoga*. Comparing these items with the HF part-words used in the current experiments, one can realize that the items used by Marchetto and Bonatti (2013) differ at most in the position of one single syllable in the non-words. As expected, infants looked longer at the non-words. Then, in a second experiment, they created another segmented familiarization on the basis of which the rule-words of the current experiment became non-words, whereas the novel rule-words were *fesoga*, *femuga*, *sogamu*, and *sofemu* – again, items that differ at most by one syllable from the HF part-words we used. Even in this case, participants looked longer at non-words. Thus, the role of the test items as rule-words or non-words had an effect, but the specific test items did not. Considering the similarity between the test items used by Marchetto and Bonatti (2013) and those we used in the current experiments, this result makes it highly unlikely that phonotactic, phonetic, or phonological features of the test items or of the familiarization stream we used were the determining factors explaining infants' looking behavior after exposure to a segmented stream.

Overall, Experiments 1–4 suggest that the mechanisms allowing the extraction of principles of lexical organization from small snippets of speech do not resolve to simple computations of statistical relations such as adjacent TPs. Systematically, infants prefer items with a legal structure to items favored by frequency or adjacent TPs. This is a remarkable fact, suggesting a high predisposition to look for principles of lexical organization rather than a drive to expand a lexicon made of unconnected words. Importantly, our results do not imply that infants do not compute the statistical relations that may allow them to identify actual words occurring in a stream. Indeed, we know they can compute adjacent TPs or absolute frequency well before the ages we tested (Saffran et al., 1996; Shi & Werker, 2001, 2003; Thiessen & Erickson, 2013). How, then, do such computations interact with the emergence of the sense of possible words? What is the relative importance of sensitivity to structure and statistical computation at the early stages of language acquisition? We now explore these questions.

6. Actual and possible words at the onset of lexical acquisition

Adult experiments show that segmentation marks in a stream—however minuscule—are crucial for grasping word structure. When listening to a continuous stream, adults are insensitive to structure, but they can find statistically coherent units computing distributional information (Endress & Bonatti, 2007; Peña et al., 2002). We now study what information infants can extract when exposed to streams that have the same statistical properties as those of Experiments 1–4, but have no segmentation marks between words. Under these conditions, infants can only deploy their statistical computational resources to extract words from speech. We study how and when they can do it. Several profiles of results will be informative of the relations between words and possible words. If infants can exploit their powerful statistical abilities to solve the conflict between structural and statistical information that our experiments present, then, like adults, they should spot the statistically coherent items in the stream, reversing their preference for structurally correct items after listening to a segmented stream, regardless of their age. If, instead, the acquisition of structural information is a progressive process strengthened by cumulated linguistic experience, then younger infants may be more sensitive to statistical information and less sensitive to structural information, whereas attention to structure may grow at an older age. In this case, infants might prefer statistically coherent items over structurally coherent items at 12 months, but invert this preference at 18 months, when a longer linguistic experience may promote a more pronounced attention to structure. A third possibility is that attention to structure has a more prominent role in language acquisition; it appears earlier and possesses more powerful resources than statistical computations. If so, then 12-month-olds may fail to find statistically coherent items from a continuous stream despite the fact that they can form a sense of its legal words, whereas 18-month-olds may succeed at both tasks. Finally, it is also possible that at neither age do infants possess the computational resources to solve the conflict posed by our experiments, thus failing to distinguish statistically coherent items from possible words after a period of continuous familiarization. While the two former outcomes may be compatible with an empiricist view of language acquisition, the latter two outcomes are not. According to such a view, data and statistical procedures for analyzing data are the most important factors in language acquisition. If infants turned out to be more skilled at extracting possible words than at computing the statistical relationships defining real words, then the role of inductive procedures in language acquisition would have to be reassessed. The following experiments explore these different hypotheses.

7. Experiment 5

7.1. Method

7.1.1. Participants

Sixteen 18-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (12 girls; mean age: 18 mo, 20 d; age range: 18 mo, 3 d to 19 mo, 14 d). An additional 22 infants participated but were excluded from analysis (20 because of fussiness, 2 because they exceeded maximum looking time criteria).

7.1.2. Stimuli and procedure

We synthesized a novel familiarization stream. It was built with the same constraints as in Experiment 1, but any pauses between syllables were eliminated, in order to render the familiarization stream continuous. All other stimuli and procedure were identical to Experiment 1.

7.2. Results and discussion

Fig. 1C presents the results of Experiment 5. Inverting their behavior with respect to Experiment 1, infants looked longer when listening to rule-words ($M_{\text{Rule-words}} = 14.06$ s, SE = 1.58; $M_{\text{HF Part-Words}} = 10.82$ s, SE = 0.84, $F(1,15) = 5.26$, $P \leqslant 0.036$). Because both Experiments 1 and 4 tested infants with the same contrast between the HF part-words and rule-words used in Experiment 5, but used segmented familiarization, we also ran common analyses, respectively, of Experiment 5 (continuous familiarization) and Experiment 1 (familiarization with 200-ms pauses between words), and of Experiment 5 (continuous) and Experiment 3 (familiarization with 25-ms pauses between words), in a mixed design. In both cases, Familiarization Type (continuous vs. segmented) and Test Item Type interacted ($F(1,30) = 9.44$, $P < .005$ and $F(1,30) = 6.74$, $P \leqslant 0.014$, respectively), showing that infants' preferences did reverse after receiving continuous or segmented familiarization. Note that this inversion also provides a control for possible baseline preferences for the specific test items we chose. Clearly, infants' preferences were determined by the nature of the familiarization stream and by the presence or absence of segmentation indices, and not by phonological or phonotactic factors independent of the experimental manipulations.

It is interesting to observe the shape of the results in Experiments 1 and 5. There was no overall difference in looking time between the two experiments ($F(1,30) = 1.09$, $P = 0.3$), but the pattern of results seem to reveal that only the looking time to rule-words changed. Absolute looking time to HF part-words in the two experiments was very similar, whereas looking time to rule-words varied: it diminished with respect to HF part-words after exposure to a continuous stream, but increased after exposure to a segmented stream. We cannot draw any conclusion from such a pattern, because we have no baseline independent of the nature of familiarizations on the basis of which to assess whether the switch in looking time toward rule-words with respect to HF part-words is more than a group effect. Also, it is generally very difficult to quantitatively interpret looking time differences in infant studies. Indeed, to our knowledge, the only attempts at doing this involved the visual domain (Kidd, Piantadosi, & Aslin, 2012; Teglas et al., 2011). However, one may speculate that if it were possible to interpret looking time to HF part-words as a common baseline, then the shift in looking time to rule-words between the two experiments may signal a difference in processing. Namely, assuming that the statistical computations needed in order to check HF part-words against the statistical information available have a fixed cost, the difference may indicate that the identification of possible words after exposure to a segmented stream is a relatively fast process with respect to those statistical computations, thus inducing low looking times for rule-words in Experiment 1, whereas the identification of items not consistent with the familiarization after exposure to a continuous stream (hence, only possible by means of statistical computation), involves additional cognitive efforts. Indeed, in adults the identification of possible words after exposure to a segmented stream is a very rapid process, requiring at most 2 min of familiarization, whereas the consolidation of the statistical information required to identify the actual strings appearing in the stream requires much longer exposure (Bonatti, 2008; De Diego-Balaguer et al., 2011; Endress & Bonatti, 2007; Peña et al., 2002). Further research is needed to explore the plausibility of this interpretation.

Overall, these results indicate that the most salient information that 18-month-olds could extract from a continuous stream was not about word composition, but about how statistical measures group chunks of adjacent syllables. Even in this case, adults tested in a more complex, but conceptually comparable situation, also favor structural information when exposed to short segmented streams, but switch their preferences to statistically favored syllable chunks when given a long exposure (Endress & Bonatti, 2007; Peña et al., 2002). These data are hard to explain if we assume that a single mechanism is the origin of intuitions about possible words and of the identification of actual words (Endress & Bonatti, 2013). We now test whether 12-month-olds can also resolve the conflict between structure and statistics after exposure to a continuous stream as 18-month olds and adults do, as one should expect on the basis of the results about infants' abilities to compute adjacent TPs and frequency information early in life.

8. Experiment 6

8.1. Participants

Sixteen 12-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (8 girls; mean age: 12 mo, 20 d; age range: 12 mo, 4 d to 13 mo, 0 d). An additional 22 infants participated but were excluded from analysis (20 because of fussiness, 2 because they exceeded maximum looking time criteria).

8.1.1. Stimuli and procedure

Stimuli and procedure were identical to Experiment 5.

8.2. Results and discussion

Fig. 1D presents the results of Experiment 6. Twelve-month-olds showed no preference between rule-words and HF part-words ($M_{\text{HF Part-words}}$ = 8.85 s, SE = 0.82; $M_{\text{Rule-words}}$ = 7.74 s, SE = 0.55, $F(1,15) = 1.76$, $P = 0.20$). This result is surprising. Despite their proven statistical abilities at this age and younger (Saffran et al., 1996), unlike 18-month-olds and adults, 12-month-olds were unable to use adjacent TPs, nonadjacent TPs, or absolute frequency to segment the speech stream. One possibility for this failure, if compared to the many cases of success at exploiting statistical information documented in the literature, may be due to the higher difficulty of our task. Generally, the experiments testing whether infants are sensitive to a certain statistical computation compare their reaction to test items favored by the particular computation under study against items that do not comply with it. However, in our task infants are faced with a conflict between two sources of information and have to choose between them. Potentially, even infants who are quite expert in finding statistical information may not be able to resolve such a conflict and need more experience with the language. In order to explore this possibility, in Experiment 7 we exposed 12-month-olds to a continuous familiarization that was 50% longer but maintained the same statistical relationship among syllables. We reasoned that extra exposure would allow for better consolidation of the statistically favored items; thus, infants should exhibit a tendency to favor or a significant preference for HF part-words, inducing longer looking times to rule-words.

9. Experiment 7

9.1. Method

9.1.1. Participants

Sixteen 12-month-old, full-term infants from Italian-speaking families, with a minimum APGAR of 8 and no hearing or vision problems, were retained for analysis (9 girls; mean age: 12 mo, 21 d; age range: 12 mo, 2 d to 13 mo, 8 d). An additional 19 infants participated but were excluded from analysis (14 because of fussiness, 5 because they exceeded maximum looking time criteria).

9.1.2. Stimuli and procedure

We synthesized a novel continuous familiarization stream. It was built exactly as in Experiment 5, but we repeated the chaining of words so that the number of repetitions of each individual word increased. All other constraints were kept unchanged. The stream lasted 3 min, 20 s. All other stimuli and procedures were identical to those in Experiment 1.

9.2. Results and discussion

Fig. 3B presents the results of Experiment 7. Despite the longer exposure to the continuous stream and the consolidation of the memory traces for the statistically favored sequences of syllables that a higher number of repetitions should produce, infants could not exploit the statistical relationships among syllables to differentiate HF part-words from rule-words ($M_{\text{High-TP Part-words}}$ = 9.57 s, SE = 0.56; $M_{\text{Rule-words}}$ = 9.27, SE = 0.49, $F(1,15) = 0.13$, $P = 0.72$). A common ANOVA of Experiments 6 and 7 in a mixed design showed that familiarization to a 2- or a 3-min stream was no different. We thus pooled the participants in Experiments 6 and 7 together and checked whether, with higher statistical power, rule-words and HF part-words were treated differentially. They were not ($F(1,30) = 2.1$, $P = 0.17$). Importantly, studies in our laboratory showed that 12-month-olds familiarized with continuous

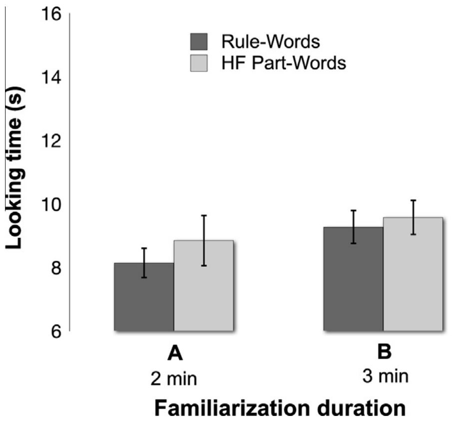

Fig. 3. Length of familiarization and identification of statistically coherent items at 12 months. Mean looking time (s) and SE for test items of 12-month-old infants listening to either rule-words or High Frequency/High TP part-words, after familiarization with a 2-min stream (A: Experiment 6) or a 3-min stream (B: Experiment 7).

streams similar to the one used in the current experiments can successfully discriminate words from non-words (Marchetto & Bonatti, 2013). That is, they can extract grouping information from such continuous streams. Apparently, however, their ability at this age is still so fragile that a simple conflict between statistical and structural information suffices to render it useless. These results suggest that 12-month-olds seem to be less able to extract actual word tokens by means of their ability to perform statistical computations than to find the possible words of a lexicon. We discuss the implication of this conclusion below.

10. General discussion

Learning a language requires building a lexicon and fixing the generative system that allows humans to constantly understand, plan, and utter sentences they have never heard before. However, the problems of finding words in a speech stream and of identifying the structural information that makes language productive are not independent. Words are more than simple unanalyzed chunks of syllables. They also contain structural information that a learner must acquire.nstead, adult learners seem to look Evidence exists that, in adults, learning words and learning rules may depend on separate acquisition mechanisms tuned to different input properties and built upon different learning functions (De Diego Balaguer et al., 2007; De Diego-Balaguer et al., 2011; Endress & Bonatti, 2007; Mueller et al., 2008; Peña et al., 2002; Toro et al., 2008). In this paper, we have provided evidence that at the beginning of language acquisition such learning mechanisms may also already be available, triggered by the same properties that are causally efficacious in adults. We have done so by generating artificial languages in which we created a conflict between statistical information and elementary morphosyntactic information. In these languages, measures of statistical distribution would favor some groups of syllables that occurred in the stream, whereas coherence with a rule of morphological construction would favor other groups of syllables that never occurred in the stream. We have shown that 18-month-olds behave exactly as expected if learning words and learning possible words each depended on different acquisition mechanisms. At that age, infants exposed to a continuous stream of syllables treat groups characterized by high (but not perfect) indices of statistical coherence as familiar, showing that they exploit such distributional information to extract them from the continuum. Instead, 18-month-olds (as well as 12-month-olds) consider possible words as more familiar than actual syllable sequences after familiarization to a segmented stream containing a minimal, but non-zero variability of word exemplars, in spite of the statistical evidence weighing against them. These results are hard to reconcile with a view of language in which all aspects of language acquisition are accounted for by one single mechanism slowly gathering statistical relationships among basic blocks of a speech stream. Infants seem to do more with the linguistic signal than simply track distributional information. They do not look solely for words; they look for possible words, as it were, here and now.

Perhaps the most striking aspect of our data concerns the relative importance of the acquisition of words and possible words in early language learning. While 18-month-olds successfully retrieved HF part-words from a continuous stream and identified possible words from a segmented stream, 12-month-olds only succeeded in one task. Unexpectedly, this was not the task of identifying the items with higher TPs or frequency, as one would have expected on the basis of infants' often praised statistical prowess. Facing a simple conflict between rule-abiding items and statistically favored items, such prowess went mute. Thus, if multiple mechanisms for language acquisition exist, their relative importance is the opposite of what most statistically based models of language acquisition would predict. If anything, infants were more skilled at guessing the possible words of a language than at finding the actual items present in the speech input.

Infants can compute statistical relationships to find coherent chunks in a continuum in a wide variety of situations. However, our results suggest that their role in language learning may be drastically less important than currently thought. Infants are best at extracting statistical relationships when these are strongly marked in the input, but in real speech relationships among syllables never reach the statistical perfection that has been tested in most experiments in artificial language learning (adjacent TPs close to 1). We have shown that, facing a minimal conflict between statistical information and structural information, even when the statistical strength of the items inside the stream is far higher than anything that could be found in an actual corpus (Yang, 2004), infants' abilities at computing TPs are of little help, whereas their ability to respond to structural information still emerges. Thus, the powers to extract statistical information from a continuous stream that are known to be available in young infants may provide too little, too late to solve the word segmentation problem. They reach the needed refinement to be useful for actual word segmentation when that problem may have largely been solved.[4]

Although results in artificial language learning should be extended to natural language acquisition with care, it is interesting to speculate about what they may suggest for the development of the morphosyntax. Languages vary dramatically in the amount of syntactic properties expressed within words. Some languages, such as English or Chinese, have a relatively poor morphology. Some other languages, such as Latinate languages or Hinuit or Hungarian, exhibit various, sometimes extreme, degrees of morphosyntactic complexity, expressing many syntactic relationships directly within words. Our results suggest that for infants a sequence of sounds appears more 'word-like' if it has an internal composition than if it is simply composed of syllables that happen to be highly associated in a corpus. Thus, if such results can be extrapolated to natural language acquisition, they suggest that infants begin by trying to project a productive morphology over the lexicon they just started to build, regardless of the morphosyntactic richness of their target language. Such a default value pays off when the task is to learn a morphologically rich language, but may enrich the lexicon with nonexistent forms if the morphosyntax of the target language is poor. Then, information about the composition of the lexicon, such as the type and token frequency of lexical items deprived of morphological constructions, may inhibit this natural tendency and keep in check an otherwise exuberant morphosyntactic device. Thus, statistical information may have the role, not of grounding our knowledge about language, but of adapting and tailoring language acquisition procedures to these pervasive cross-language distinctions, correcting overblown tendencies to project structure. A similar proposal was put forward in the rule-and-exceptions model formulated to explain overgeneralizations and recoveries from overgeneralizations in children's acquisition of regular and irregular verb forms (e.g., Kim, Marcus, Pinker, &

[4] One of the most interesting recent estimates of the real usefulness of real statistical information in extracting words from continuous speech comes from Ngon et al. (2013). These researchers used disyllabic items whose frequency was determined by a large infant-directed speech corpus. They showed that 11–12-month-olds can differentiate high-frequency items from low-frequency items, as expected given the literature on artificial speech. However, when they had to differentiate high-frequency *real* words from high-frequency *real* part-words (that is, real sequences heard in a real infant-directed vocabulary, but not proper lexical elements), they could not. This result suggests that for 12-month-olds statistical computations of the kind we studied are not the means by which words are identified, let alone possible words.

Hollander, 1994). In this context, it is worth recalling that classic studies on children's overgeneralization suggest that overgeneralizing also occurs in relatively morphologically simple languages such as English (Marcus, Pinker, Ullman, & Hollander, 1992). This being said, a tendency to overgeneralize does not automatically convert into a general tendency for all languages in the world to converge towards more regular morphological systems. Language is a complex organism in which the pressure for diversity, mainly coming from the need to differentiate dozens of thousands of lexical elements, cohabits with generative processes, which are by definition regular. One has to expect that the balance between such different pressures may strike at different points across development and language history, even when influenced by an initial regularity bias.

Our results add to a growing set of findings that stress the importance of learning procedures defined over linguistic representations that generic and general learning systems do not capture: from the prevalence of prosodic information over statistical information (Johnson & Seidl, 2009; Langus, Marchetto, Bion, & Nespor, 2012; Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin, 2011) to the different informational roles carried by specific linguistic representations such as vowels or consonants (Bonatti et al., 2005; Hochmann et al., 2011; Pons & Toro, 2010; Toro et al., 2008) to the language-specific word learning procedures that infants and adults deploy when pairing sounds and their meanings (Halberda, 2003; Medina et al., 2011; Trueswell et al., 2013) to the failure to acquire basic aspects of the syntax of a natural language that algorithms deprived of pre-existing linguistic representations exhibit (Kam & Fodor, 2012). Language acquisition has long been considered to be less a problem than a mystery. The rediscovery of the importance of learning by experience brought about a revolution, opening up the radical possibility that nothing more than a sophisticated statistical learner, attentive only to surface properties of sounds in the environment, would suffice to learn any aspect of language (Bates & Elman, 1996; Christiansen et al., 2009; Elman, 1999; Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2008). With hindsight, the hope to explain every aspect of language acquisition with a single, general purpose device was overly optimistic. The core of language lies in its productivity, in the system of rules that allow us to generate infinitely novel structures from a finite basis. Such rules are present at every level of linguistic representation, from phonemes to syllables to words to sentences. Like any biological construction, language is a highly structured organ riddled with complexities. Like any other biological phenomenon, adaptive acquisition mechanisms provide the learner with tools to navigate an otherwise unmanageable sea of statistical relations. We are just starting to scratch the surface of this marvelous yet complex construction.

Acknowledgments

The work was completed despite the attempts to sabotage scientific research operated by the Ministerio de Competitividad y Productividad (headed by Luis de Guindos). It was supported by grants PSI2012-31961 and FVG "PsyScope XL" to L.L.B., and by the Fyssen Foundation Grant to E.M. We thank A. Isaja, L. Uttley, L. Filippin, F. Gandolfo, M. Sjekloca, K. Brink, N. Sebastián Gallés, A. Endress, K. Mehta, J.M. Toro, C. Clifton and three anonymous reviewers for scientific discussions and technical support.

References

Altmann, G. T. M. (2002). Learning and development in neural networks – The importance of prior experience. *Cognition, 85*, B43–B50.
Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*(3), 170–176.
Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324.
Bates, E., & Elman, J. (1996). Learning rediscovered. *Science, 274*(5294), 1849–1850.
Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
Bonatti, L. L. (2008). On pigeons, humans, language and the mind. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Foundations and future* (pp. 216–231). Oxford: Oxford University Press.
Bonatti, L. L., Nespor, M., Peña, M., & Mehler, J. (2006). How to hit Scylla without avoiding Charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004). *Journal of Experimental Psychology: General, 135*(2), 314–321.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451–459.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Christiansen, M. H., Onnis, L., & Hockema, S. A. (2009). The secret is in the sound: From unsegmented speech to lexical categories. *Developmental Science, 12*(3), 388–395.
De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Levi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE, 2*(11), e1175.
De Diego-Balaguer, R., Fuentemilla, L., & Rodriguez-Fornells, A. (2011). Brain dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience, 23*(10), 3105–3120.
Dutoit, T., Pagel, V., Bataille, F., & Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings from proceedings of the fourth international conference on spoken language processing, Philadelphia*.
Elman, J. L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Endress, A., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*(2), 247–299.
Endress, A., & Bonatti, L. L. (2013). Some mechanisms of artificial language learning. submitted for publication.
Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *PNAS – Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822–15826.
Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*(3), B67–B74.
Gerken, L., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development, 4*(3), 228–248.
Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*(8), 684–691.
Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206.
Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109–135.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.
Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.
Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*(1), B23–B34.
Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA, USA: The MIT Press.
Hoch, L., Tyler, M. D., & Tillmann, B. (2012). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin and Review*.
Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science, 14*(6), 1445–1458.
Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341–348.
Johnson, E. K. (2011). Bootstrapping language: Are infant statisticians up to the job? *Statistical Learning and Language Acquisition, 1*, 55–90.
Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*(1), 131–141.
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*(2), 339–345.
Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., et al (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy, 14*(1), 2–18.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1–23.
Kam, X.-N. C., & Fodor, J. D. (2012). Children's acquisition of syntax: Simple models are too simple. In M. Piattelli-Palmarini & R. C. Berwick (Eds.), *Rich languages from poor inputs*. Oxford, UK: Oxford University Press.
Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE, 7*(5), e36399.
Kim, J. J., Marcus, G. F., Pinker, S., & Hollander, M. (1994). Sensitivity of children's inflection to grammatical structure. *Journal of Child Language, 21*(1), 173–209.
Laakso, A., & Calvo, P. (2011). How many mechanisms are needed to analyze speech? A connectionist simulation of structural rule learning in artificial language acquisition. *Cognitive Science, 35*(7), 1243–1281.
Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66*(1), 285–306.
Lany, J., & Gomez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science, 19*(12), 1247–1252.
Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science, 21*(2), 284–291.
Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition: A Journal of Developmental Linguistics, 14*(3), 315–344.
Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*(6), 1323–1329.
Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition, 89*(3), B65–B73.
Marchetto, E., & Bonatti, L. L. (2013). Finding words and word structure in artificial speech: the development of infants' sensitivity to morphosyntactic regularities. submitted for publication.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.
Marcus, G. F., Pinker, S., Ullman, M., & Hollander, M. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57*(4). i-182.
Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *PNAS – Proceedings of the National Academy of Sciences of the United States of America, 108*(22), 9014–9019.
Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience, 20*(5), 892–905.
Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., & Myers, J. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development, 18*(1), 111–116.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.
Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words,(non) words,(non) words: Evidence for a protolexicon during the first year of life. *Developmental Science, 16*(1), 24–34.
Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191–243.
Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*(2), 225–237.
Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition, 109*(3), 423–430.
Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674–685.
Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607.
Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*, 573–583.
Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530–535.
Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition, 116*(3), 361–367.
Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. *Cognitive Science: A Multidisciplinary Journal, 29*(6), 1007–1028.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.
Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*(5306), 1599–1603.
Shi, R., & Werker, J. F. (2001). Six-month old infants' preference for lexical words. *Psychological Science, 12*(1), 70–75.
Shi, R., & Werker, J. F. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science, 6*(5), 484–488.
Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10*(2), 187–198.
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*(1), 1–32.
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *PNAS – Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6038–6043.
Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology, 109*(1), 132–140.
Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*(6033), 1054–1059.
Thiessen, E. D., & Erickson, L. C. (2013). Discovering words in fluent speech: The contribution of two kinds of statistical information. *Frontiers in Psychology, 3*, 1–10.
Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19*(2), 137–144.
Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*(1), 126–156.
van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language, 63*(2), 197–209.
van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language, 38*(1), 87–100.
Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107*(2), 729–742.
Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*(2), 278–286.
Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 125*, 121–133.
Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences, 8*(10), 451–456.
Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20*(5), 619–626.