Cognitive Science 34 (2010) 255–285
Copyright © 2009 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01074.x

Exploiting Multiple Sources of Information in Learning an Artificial Language: Human Data and Modeling

Pierre Perruchet,[a] Barbara Tillmann[b]

[a]*CNRS-UMR 5022, University of Bourgogne*
[b]*CNRS-UMR 5020, University of Lyon 1*

Received 4 April 2008; received in revised form 17 July 2009; accepted 22 July 2009

Abstract

This study investigates the joint influences of three factors on the discovery of new word-like units in a continuous artificial speech stream: the statistical structure of the ongoing input, the initial word-likeness of parts of the speech flow, and the contextual information provided by the earlier emergence of other word-like units. Results of an experiment conducted with adult participants show that these sources of information have strong and interactive influences on word discovery. The authors then examine the ability of different models of word segmentation to account for these results. PARSER (Perruchet & Vinter, 1998) is compared to the view that word segmentation relies on the exploitation of transitional probabilities between successive syllables, and with the models based on the Minimum Description Length principle, such as INCDROP. The authors submit arguments suggesting that PARSER has the advantage of accounting for the whole pattern of data without ad-hoc modifications, while relying exclusively on general-purpose learning principles. This study strengthens the growing notion that nonspecific cognitive processes, mainly based on associative learning and memory principles, are able to account for a larger part of early language acquisition than previously assumed.

Keywords: Psychology; Attention; Language acquisition; Learning; Memory

1. Introduction

Language acquisition initially proceeds from auditory input, and linguistic utterances usually consist of sentences linking several words without clear acoustical boundaries. The question thus arises: How do infants discover the words of their language? The seminal

Correspondence should be sent to Pierre Perruchet, CNRS-UMR 5022, Université de Bourgogne, LEAD/CNRS, Pole AAFE, Esplanade Erasme, 21000 Dijon, France. E-mail: pierre.perruchet@u-bourgogne.fr

studies by Saffran and collaborators (e.g., Saffran, Newport, & Aslin, 1996) have allowed a major advance on this issue. These studies have shown that infants, children, and adults can extract the word-like units (hereafter: the words) from an artificial language in which these units have been concatenated without any phonological or prosodic markers. This attests to the fact that listeners are able to exploit the statistical information available in language. For instance, participants would be able to find the word *befoki* from a sequence such as ''…*dubefokita*…,'' because, when assessed from a sizeable corpus, the degree of cohesiveness between word internal syllables (e.g., between *fo* and *ki*) is stronger than the degree of cohesiveness between syllables spanning word boundaries (e.g., between *du* and *be*).

As has often been emphasized, however, the fact that statistical cues turn out to be sufficient to extract words from artificial languages does not mean that they are exclusive factors in natural language acquisition. Many other sources of information have been documented so far (for a review: Jusczyk, 1997). The first objective of this study is to further document the way different factors interact when they are present together in an artificial language. The second objective is to examine the ability of a specific word segmentation model, PARSER (Perruchet & Vinter, 1998), to simulate the influence of these factors. A major interest of PARSER is its ability to account for the exploitation of the statistical structure of the language while relying exclusively on general-purpose learning principles. If PARSER successfully simulates the effect of additional sources of information and their interplay, this would strengthen the growing idea that nonspecific cognitive processes, mainly based on associative learning and memory principles, are able to account for a substantial part of early language acquisition.

In the following, we first describe the two additional sources of information that have been manipulated in this study. Then we present the principles underlying PARSER, and we outline how PARSER could, in principle, account for the exploitation of all three sources of information. Of course, several other models have been proposed for word segmentation (for a review, see Brent, 1999), and in addition, general memory models, such as MINERVA 2 (Hintzman, 1986) or ETAM (Logan, 2002) could also be examined for their ability to account for word segmentation in general and our findings in particular. As an exhaustive examination of all models would exceed the scope of this study, PARSER will be compared in the final discussion with some of the most common approaches, notably with the view that word segmentation relies on the exploitation of transitional probabilities (Aslin, Saffran, & Newport, 1998; Saffran et al., 1996) or on more sophisticated measures of association between successive syllables (Swingley, 2005), and with the models based on the Minimum Description Length (MDL) principle, such as the INCDROP model (e.g., Brent, 1996; Brent & Cartwright, 1996).

1.1. The initial word-likeness

Anyone exposed to an unknown speech flow may observe that some parts of it have a stronger tendency to be perceived as a word than other parts. In the following, we refer to the propensity of a given set of sounds (due to its intrinsic properties) to be considered as a word by a given listener as its *initial word-likeness* (IWL). The degree of IWL may be due to a number of factors. Some of these factors are certainly universal, gestalt-like cues for the formation of perceptual units, while others may be more specifically linked to a given language, and hence dependent on the learners' experience with their mother tongue. The role of phonological and prosodic features, such as lexical stress placement, on word discovery has been well documented (e.g., Creel, Tanenhaus, & Aslin, 2006; Curtin, Mintz, & Christiansen, 2005; Thiessen & Saffran, 2007). The question of how statistical and phonological or prosodic cues combine has been investigated in experimental studies in which phonological and prosodic cues either helped or hindered the discovery of artificial words in a continuous speech flow. These studies (e.g., Creel et al., 2006; Onnis, Monaghan, Chater, & Richmond, 2005; Shukla, Nespor, & Mehler, 2007; Tyler, Perruchet, & Cutler, 2006) have shown that performance in a word segmentation test improved in the former case and strongly decreased (and potentially dropped at chance level) in the latter case. Other studies have assessed the respective influences of these factors on word segmentation as a function of age. For instance, Johnson and Jusczyk (2001) reported that prosodic factors override statistics in 8-month-old infants, whereas Thiessen and Saffran (2003) reported a prevalence of statistics over prosody in 6-month-old infants. Other factors may also contribute to the relative IWL of a given set of sounds. For instance, the inclination of learners to consider a given part of a new speech flow as a word may also depend on its similarity with one or several words from the learners' native language (even if Magnuson, Tanenhaus, Aslin, & Dahan, 2003 have shown that this factor could be relatively unimportant, at least in the early phases of training).

1.2. The role of context

The probability of discovering the words of a new language also depends on the context in which they occur. The direct exposure to a speech flow composed of entirely new words, as in a laboratory setting, is certainly an unusual experience in children's everyday life. Children necessarily acquire some words before others. A mechanism at play in natural language acquisition could be the exploitation of known words to discover new words. To borrow an example given by Dahan and Brent (1999): ''If *look* is recognized as a familiar unit in the utterance *Lookhere!* then *look* will tend to be segmented out and the remaining contiguous stretch, *here*, will be inferred as a new unit'' (p. 165). To provide an experimental evidence of this phenomenon, Dahan and Brent (Experiment 1) exposed participants to short nonsense utterances (e.g., either a two-syllable word *ab* or a three-sylable word *abc*) and to long utterances beginning with the short utterances (e.g., a five-syllable sequence *abcde*). Then participants were presented with items such as *cde* and *de*, and they were asked to decide whether this item was a word from the artificial language they had just heard. Participants classified a test item as being a word more frequently when this item was the remainder of the long utterance after extraction of the short utterance (i.e., *cde* when the short utterance was *ab*, and *de* when the short utterance was *abc*). Bortfeld, Morgan, Golinkoff, and Rathbun (2005) demonstrated the same ability in 6-month-old infants. In this case, the first part of the utterance was a highly familiar word, such as *Mommy*. It has been suggested that such lexically driven segmentation could progressively supersede prosodic and phonological cues during language development (e.g., Matthys, White, & Melhorn, 2005, p. 493).

1.3. PARSER and the exploitation of statistical information

Before examining how PARSER can account for the exploitation of the two sources of information described above, we first need to examine the principles underlying the model, and why these principles make it able to exploit the statistical structure of a language. Perruchet and Vinter (1998, 2002) have proposed that word extraction happens as a direct consequence of the organization of the cognitive system. They characterized this organization as the interplay of two interrelated principles. The first principle stipulates that perception shapes internal representations. This means that the primitives that are perceived within one attentional focus as a consequence of their experienced spatial or temporal proximity (i.e., they are perceived as a chunk) become the constituents of one new representational unit. The future of this provisional unit, they argued, depends on ubiquitous laws of associative learning and memory. If the association between the primitives that form a provisional unit is not strong enough in the language, this representation rapidly vanishes, as a consequence of both natural decay and interference with the processing of similar material. However, if the degree of cohesiveness between the primitives is sufficient, the internal representation is progressively strengthened.

The second principle is that internal representations guide perception. Perception involves an active coding of the incoming information constrained by the perceiver's knowledge. Internal representations serve as perceptual primitives. Because the representational landscape changes with increased experience in a domain, perception, and notably the composition and the size of the perceived chunks, also evolves. The resulting picture is that perception builds the internal representations which, in turn, guide further perception, hence leading to the self-organization of the mind (Perruchet & Vinter, 2002). The mutual dependence of perception and internal representations is in line with a developmental principle initially described by Piaget's concepts of assimilation and accommodation (e.g., Piaget, 1985). Most current theories of development, although they use different terminology, also rely on the constructive interplay between assimilation-like and accommodation-like processes (e.g., Case, 1993; Fischer & Granott, 1995; Karmiloff-Smith, 1992). A similar view, which contrasts with the claim that perception is driven by a fixed repertoire of primitive features, has been cogently documented by Schyns, Golstone, and Thibaut (1998) for visual perception.

These principles have been exploited in PARSER, a chunk-based computational model devised to discover words from a nonsegmented speech flow (Perruchet & Vinter, 1998). How does PARSER work? Based on the phenomenon that, in humans, attentional coding[1] of the ingoing information naturally segments the material into disjunctive parts, the model is provided online with a succession of candidate units, such as *foki* and *dube* in our example above, some of them relevant to the structure of the language and others irrelevant. According to the first principle described above, an internal representation that matches a percept is reinforced in the model if its components are cohesive and occur repeatedly in the input.

This means that a word or a part of a word (e.g., *foki*, because *befoki* is a word) are more likely to create a long-lasting internal representation than between-word segments (e.g., *dube*, because in our example, *dube* spans over a word boundary). The relevant units emerge through a selection process based on forgetting. Forgetting due to both decay and interference[2] leads to the selection of the most cohesive parts among all parts generated by the initial, presumably mostly irrelevant, chunking of the material. The second principle described above ensures the convergence of this process toward an optimal parsing solution. The fact that perception is guided by internal representations allows the system to build representations of words whose components could hardly be perceived in one attentional focus if perception were driven only by the initial primitives in the language. Also, once internal representations providing an appropriate coding of the input have been built, an endless generation of new candidate units is avoided. Previous studies have shown that PARSER is able to exploit the statistical cues available in the input (Perruchet & Desaulty, 2008; Perruchet & Vinter, 1998; Perruchet, Vinter, Pacteau, & Gallego, 2002). However, up to now, PARSER has only been applied to cases in which statistical information provided the sole segmentation cue available in the language.

1.4. How can PARSER account for the exploitation of additional segmentation cues?

How can the influence of IWL be implemented in PARSER? Although PARSER has never been made sensitive to IWL in prior simulations, the model has a natural way to integrate this feature. In the original version of the model (Perruchet & Vinter, 1998), the initial formation of candidate units (which are subsequently selected by the action of decay and interference) was randomly determined. More precisely, a given unit might comprise one, two, or three perceptual primitives. However, Perruchet and Vinter noted that selecting at random the length of the provisional units was only a convenient way of simulating multiple determinants of the initial segmentation, including ''prior experience with another language'' and ''the relative perceptual saliency of the components of the signal'' (p. 249). This is a direct consequence of the fact that a chunk corresponds to one attentional focus, and that the content of attentional processing is determined by the interaction between the perceiver's knowledge and the properties of the input. Implementing IWL into PARSER is thus straightforward. In the simulations below, the initial selection of the candidate units will be biased in such a way that instead of being randomly drawn within a given length range, the candidate units will be selected (within the same range) as a function of their relative IWL, which had been assessed with a previous experiment on human participants (Perruchet, Tyler, Galland, & Peereman, 2004).

Regarding now the exploitation of the surrounding context, PARSER, in principle, should be efficient without any modification. This is because the coding of the incoming information is constrained by the perceiver's internal representations. More precisely, the disjunctive partition of the sensory input that PARSER makes throughout its exposure to the language is guided by the current perceptual primitives of the model. If the perceptual primitives are the syllables, as it is postulated at the outset of training, this means that the cutting edge between two candidate units will fall always at a syllable boundary, and not, say, between two syllable-internal phonemes. Now, in PARSER, the perceptual primitives evolve throughout training, passing (ideally) from the syllable to the words. When a multi-syllable unit becomes a new perceptual primitive, this provides a natural constraint on the possible partitions of the speech flow. If PARSER is exposed to *abcde* while *ab* has been previously built as a perceptual primitive within the model, the following percept necessarily begins by *c* (it can be *c*, *cd*, or *cde*, depending on random selection and on the available perceptual primitives), hence increasing the probability of discovering *cde* (if *ab* is not a perceptual primitive, provisional units such as *abc* and *de*, or *a*, *bcd*, and *e*, could be created).

1.5. The present study

In summary, it is possible to distinguish at least three categories of factors involved in word segmentation of an artificial speech flow: the statistical cues (i.e., the dependency relations between the primitives of the artificial speech flow), the IWL (i.e., the fact that the intrinsic properties of a new sound sequence make it more or less likely to be considered as a word of the language, irrespective of the origins of these initial biases), and the contextual cues (i.e., the information provided by the speech flow surrounding a given sound sequence). Our aim was to investigate participants' behavior in an experimental situation allowing us to explore the interplay between these three sources of information when all of them are made jointly available (although not necessarily in a congruent way), and to compare the ability of PARSER and a few other common approaches to simulate the observed results.

Our situation was quite similar to the situation introduced by Saffran and collaborators (e.g., Saffran et al., 1996). Participants had to listen to an artificial language composed of six trisyllabic artificial words, randomly concatenated without any pauses. The main difference with regard to the standard situation was that the IWL was systematically manipulated for three of the six artificial words. For one group of participants (Group IWL+), these three words, when heard in a continuous speech stream, were spontaneously perceived as words more often than trisyllabic units spanning word boundaries (referred to hereafter as part-words), which may comprise either the last two syllables of a word and the first syllable of the next word, or the last syllable of a word and the first two syllables of the next one. For a second group of participants (Group IWL−), the words were composed from the same set of syllables, but these syllables were arranged in such a way that the resulting words were spontaneously perceived as such *less* often than the resulting part-words. In the test phase, participants were presented with two-alternative forced choice paradigm (a word and a part-word) and had to select the syllable set forming a word in the previously heard syllable stream.

Our hypothesis was that positively biased words should be learned more quickly than negatively biased words. This prediction may be thought of as self-evident, because if some chunks of syllables are spontaneously detected as words in a continuous speech stream, then they might also be preferred when played in isolation during the forced-choice test, hence improving the word/part-word discrimination score. However, this was not our point. We hypothesized that IWL should affect *learning*, construed as the exploitation of the structure of the ongoing speech flow, beyond its *direct* effects on performance in the test. To control for possible direct effects of IWL, the same word/part-word forced-choice test was performed twice. The first test was primarily devised to serve as a baseline to assess learning. It occurred after a very limited exposure to the language (a small amount of language exposure, instead of no exposure at all, was provided in order to give sense to the instructions, which referred to the speech flow that had just been played). The second test occurred at the end of the experiment, as usual. The effects of IWL should be captured in the first test, and hence the change in performance between the two tests should provide a reliable measure of learning—and, most crucially, reveal how learning is affected by IWL. Note that in most studies, the direct effect of perceptual biases is controlled by withdrawing those biases from the test items (Johnson & Jusczyk, 2001; Shukla et al., 2007; Thiessen & Saffran, 2003). However, as a consequence, test items differ from study items in their perceptual properties and this change in material might influence test performance, notably leading to underestimation of learning (see Shukla et al., 2007, Experiments 1 vs. 3).

In addition to the three biased words, three other trisyllabic words were composed by randomly concatenating nine syllables. Different combinations of syllables were used for each participant. This procedure ensured that, on average, the IWL of these words and the IWL of the part-words that are generated by the concatenation of these words did not differ. Performance on these words should indicate how much participants are able to exploit their growing knowledge of other words to guide further learning. The unbiased words, although identical with regard to their statistical properties and their IWL, should be learned more quickly in the Group IWL+ than in the Group IWL− because of the role of context, as explained above (assuming our first hypothesis correct, i.e., that participants were sensitive to the IWL of the biased words).

2. Experiment

2.1. Method

2.1.1. Participants

A total of 40 undergraduate students from the University of Bourgogne in Dijon, France, participated in the experiment in partial fulfillment of a course requirement. All subjects were native French speakers. Participants were randomly assigned to one of the two experimental groups (IWL+ and IWL−).

2.1.2. Materials

The language was composed of six trisyllabic words, half of them differing between groups. The selection of the words that differed between groups (the biased words) was based on the results of an earlier experiment (Perruchet et al., 2004, Experiment 1). In this experiment, 12 participants (only the group ‘‘no-gap’’ is relevant here) were exposed to an artificial language composed of 27 different words, obtained by the exhaustive combination of three first syllables (*pu*, *be*, and *ta*), three medial syllables (*li*, *ra*, and *fo*), and three last syllables (*ki*, *ga*, and *du*). These words were ordered in such a way that two consecutive words did not share any syllable. A striking property of this language is that it is ''statistically flat.'' This means that all sets of three consecutive syllables, whether they are words or part-words, have exactly the same frequency. Likewise, the intra-word transitional probabilities are exactly the same as the transitional probabilities between syllables spanning a word boundary. After having heard the 27-word sequence (a different order was generated for each participant), participants were informed that this language was composed of words. They were told that other similar samples of this language would be played, and that, while listening, they would have to write down on a sheet of paper the words that they perceived. They then listened to four successive 40-s long samples of the language. Each sample was composed of one instance of each different word. Progressive fades in and out were applied to the first and last 5 s of each sample in order to avoid word boundary cues.

For the Group IWL+ of the present experiment, we selected three trisyllabic words that were often written down, while sharing no common syllable. These words were *befoki*, *pulidu*, and *taraga*. For the Group IWL−, half of the participants (those with an odd number) were exposed to *fokipu*, *liduta*, and *ragabe*, and the other half (those with an even number) to *kipuli*, *dutara*, and *gabefo* (none of these words appeared in participants' production in Perruchet et al., 2004). The words for the Group IWL+ served as part-words for the Group IWL− and vice versa. The perceptual biases that are captured here have certainly mixed origins. An important factor may be that all of the first and final syllables began with a stop consonant, while all of the medial syllables began with a continuant consonant (Onnis et al., 2005; Perruchet et al., 2004; Seidenberg, MacDonald, & Saffran, 2002). However, other factors, such as the phonological similarity with a French word, may have played a role as well. For the present study, our only objective was to distinguish two categories of units as a function of their relative IWL, without teasing apart the presumably intricate influences underlying this property.

The other three trisyllabic words (the unbiased words) were composed by randomly concatenating nine syllables (*mɑ̃, ne, so, vy, ʒɛ, dɛ̃, ti, lɔ̃, kø*). Different words were used for each participant, hence making it quite unlikely that the material contains some residual biases. However, as a further control, participants from the groups IWL+ and IWL− were coupled, in such a way that Participant #n from the Group IWL+ was exposed to the same unbiased words as Participant #n from the Group IWL−. As a consequence, any between-groups difference in performance on unbiased words ought to be attributed to an influence of biased words, rather than to an uneven selection of unbiased words.

The speech was synthesized through the MBROLA (Multiband Resynthesis Overlap Add) speech synthesizer (http://tcts.fpms.ac.be/synthesis/; Dutoit, Pagel, Pierret, Bataille, & Van Der Vrecken, 1996) with the FR2 diphone database. The mean syllable duration was 232 ms. The resulting WAV file was modified using CoolEdit. Progressive fades in and out were applied to the first and last 5 s of each part of stream to avoid word boundary cues. The speech stream was played through headphones connected to a personal computer using CoolEdit.

2.1.3. Procedure

Participants were told that they would listen to an imaginary language. They were asked to avoid engaging in analytic, problem-solving processes. The initial phase of familiarization to the language lasted about 40 s. Each of the six words occurred 10 times. The words were pseudo-randomly ordered for each participant, without immediate repetition. After this initial phase, participants were told that they would be presented with pairs of items, and that they would have to judge, for each pair, which item seemed more like a word of the imaginary language. There were 36 pairs of items, with the two items of each pair being separated by a 500-ms silent interval. Eighteen pairs involved trisyllabic items. Nine pairs of trisyllabic items involved the biased words, and nine pairs involved the unbiased words. For the test pairs with biased words, the words for the Group IWL+ were the part-words for the Group IWL− and vice-versa. For the test pairs with unbiased words, words and part-words were the same for each couple of participants. The remaining 18 pairs involved bisyllabic items. They contrasted word internal bigrams to bigrams spanning over word-boundaries. These bisyllabic items were introduced in order to avoid that participants learned from the initial test that the language was composed of trisyllabic words, hence potentially orienting their perception during the following exposure phase (for further details about the construction of the material, and the exhaustive list of test items, see Appendix A). The order of the items within a pair and the order of the pairs in the test sequence were randomized for each pair of coupled participants.

After the first test, participants were told that the imaginary language would be played again for about 8 min. Each of the six words occurred 120 times. At the end of this phase, participants were submitted to the same test as previously (except the order of the items within a pair, and the order of the pairs in the test sequence, which were randomized for each pair of coupled participants).

2.2. Results

An ANOVA was performed with Group (IWL+ vs. IWL−) as a between-subjects factor and Tests (initial vs. final), Word status (biased vs. unbiased), and Length (tri- vs. bi-syllabic) as within-subjects factors. Because the length of the items elicited neither main nor interactive effects in the ANOVAS (all ps > .05), the data displayed in Fig. 1 were pooled over trisyllabic and bisyllabic test items. As expected, the Group IWL+ outperformed the Group IWL−, $M = 0.715$, $SD = 0.112$ versus $M = 0.547$, $SD = 0.129$, $F(1, 38) = 19.27$, $p < .001$, $\eta_p^2 = 0.336$, and final performance was better than initial performance, $M = 0.689$, $SD = 0.177$ versus $M = 0.574$, $SD = 0.142$, $F(1, 38) = 35.36$, $p < .001$, $\eta_p^2 = 0.482$. The interaction between the factors Group and Test was significant, $F(1, 38) = 5.59$, $p = .023$, $\eta_p^2 = 0.128$, indicating that language exposure was more beneficial to the Group IWL+ than to the Group IWL−. More crucially for our concern, this interaction was not modulated by whether the biased items or the unbiased items were considered, as indicated by the nonsignificant Group × Test × Word Status interaction, $F(1, 38) = 0.51$, $p = .478$, $\eta_p^2 = 0.013$. Responses on unbiased items were better in the Group IWL+ than in the Group IWL−, although these items were identical in the two groups. This difference did not reach

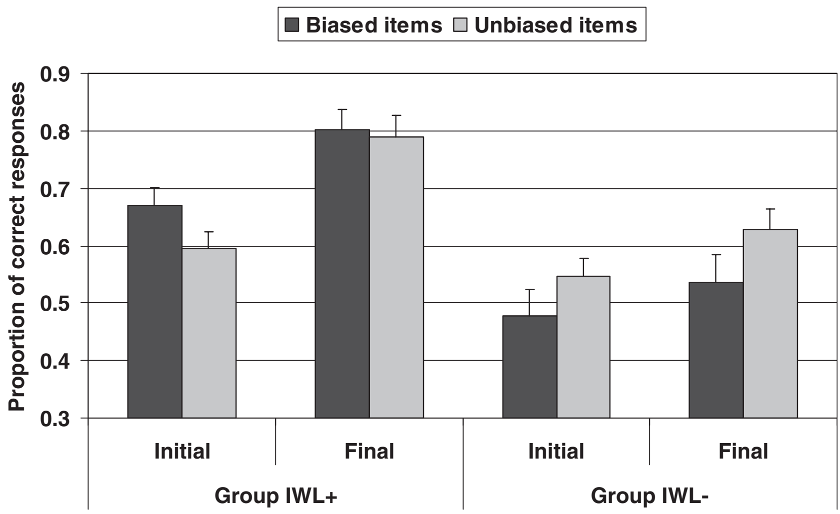


Fig. 1. Proportion of correct responses in the initial and the final tests, as a function of Groups, for biased and unbiased items. The biased items were positively biased in the Group IWL+, and negatively biased in the Group IWL−. Unbiased items were identical for the two Groups. Error bars represent standard errors.

significance during the initial test, $M = 0.594$, $SD = 0.131$ versus $M = 0.547$, $SD = 0.138$, respectively, $F(1, 38) = 1.23$, $p = .273$, $\eta_p^2 = 0.031$, but was highly significant during the final test, $M = 0.789$, $SD = 0.167$ versus $M = 0.628$, $SD = 0.157$, respectively, $F(1, 38) = 9.88$, $p = .003$, $\eta_p^2 = 0.206$, hence generating a significant Group × Test interaction for unbiased items, $F(1, 38) = 5.02$, $p = .031$, $\eta_p^2 = 0.117$).

All the other main effects and interactions of the ANOVA were non significant, except the two-way interaction between Group and Word Status, $F(1, 38) = 5.65$, $p = .023$, $\eta_p^2 = 0.129$. For the Group IWL+, correct responses (pooled over initial and final tests) were more numerous for the biased items ($M = 0.739$, $SD = 0.126$) than for the unbiased items ($M = 0.692$, $SD = 0.124$), whereas for the Group IWL−, correct responses were more numerous for the unbiased items ($M = 0.587$, $SD = 0.127$) than for the biased items ($M = 0.507$, $SD = 0.193$). However, in each case, these differences were marginally significant at best, $F(1, 19) = 3.27$, $p = .087$, $\eta_p^2 = 0.147$; $F(1, 19) = 2.94$, $p = .103$, $\eta_p^2 = 0.134$, respectively.

Did learning actually occur in the Group IWL−? When data were restricted to this group, performance in the final test ($M = 0.582$, $SD = 0.137$) was still better than performance in the initial test, $M = 0.512$, $SD = 0.146$, $F(1, 19) = 7.26$, $p = .014$, $\eta_p^2 = 0.276$, attesting for learning. The level of learning was comparable for biased and unbiased items, as shown by the nonsignificant interaction between tests and items, $F(1, 19) = 0.26$, $p = .612$, $\eta_p^2 = 0.014$. These results suggest that negatively biased items were learned. However, it is possible that the discovery of negatively biased items in the Group IWL− was a consequence of the segmentation of the unbiased items, in the same way as the discovery of the unbiased items may be a consequence of the segmentation of biased items in the Group

IWL+. A conservative conclusion is that statistical learning for negatively biased items is not yet demonstrated, and that further studies involving only negatively biased items are needed to settle the issue.

3. Simulations

3.2. Method

In PARSER (Perruchet & Vinter, 1998), the primitives that are perceived within one attentional focus are assumed to become the constituents of a new representational chunk, which can be thought of as a provisional candidate word (or part of a word). If those primitives are not cohesive enough, the representation of the corresponding chunk rapidly vanishes, as a consequence of both decay and interference due to the processing of units sharing the same primitives. However, if the primitives composing a chunk are strongly associated, this chunk is progressively strengthened and serves to guide further perception. The detailed algorithm is described in Appendix B.

As noted in the Section 1, PARSER has never been made sensitive to IWL in prior simulations, but the model is endowed with a natural slot to do so, namely the initial selection of the candidate units from the speech stream. The general principle is that these units are attentionally driven and, as a consequence, are restricted to a small number of primitives. However, some indeterminacy remains. Perruchet and Vinter (1998) noted that the selected part of the sensory input probably depends on a number of features, such as the perceptual saliency of the material. Because these factors have no reason to be correlated with word boundaries, at least for artificial languages, their action was simulated by drawing at random the number of perceptual primitives embedded in each provisional units within a (1–3) range. Implementing a perceptual bias in PARSER was straightforward: The rationale consisted in restoring the role of perceptual factors wherever it was artificially eliminated by a randomly based selection in the original instantiation of the model.

It seems reasonable to assume that if a given set of syllables is spontaneously perceived as a perceptual unit, it will be processed as a whole, insofar as this does not exceed the processing capacity of the system. This hypothesis was implemented in PARSER to simulate the effect of strong IWL (as shown by Perruchet et al., 2004). If, at a given processing step, a positively biased unit was among the possible choices (i.e., the choices respecting the general constraints of the model), then this unit was selected. This amounts to saying that the sole effect of perceptual biases was to change the randomly drawn number of perceptual primitives composing the current percept (within the standard [1–3] range) whenever this change leads to select a positively biased unit. Note that the very same procedure was used for the Group IWL+ and the Group IWL−. What made the difference between the groups was that the positively biased unit was a word of the language in the first case, and a part-word in the second case. Because the algorithm of PARSER was left unchanged, the reader may refer to Perruchet and Vinter (1998) for further precisions about the characteristics of the model.

For the sake of comparison with the experiment above, 20 simulations were run for each of the two levels of decay explored in this study, as discussed in more detail later. All the randomly generated features, including the order in which the words of the languages were displayed, differed between simulations. The number of items used for the simulations was the same as for the behavioral experiment, namely: Each of the six words occurred 10 times during the first phase and 120 times during the second phase.

3.2. Data analysis

Throughout training, PARSER is creating chunks, and each chunk is attributed a weight, which may be conceived of as its strength in memory. The chunks ideally match the words of the language when learning is at asymptote, but depart from those words whenever training is still in progress. Assessing the performance of the model in a forced-choice test identical to the test run by the participants needs a model of performance, which is independent from PARSER itself. In the following, a response was generated for each word/part-word pair, based on the ratio between the weights of the word and the part-word. Assuming that the word has a weight of 3 and the part-word has a weight of 2 for a specific pair, a response was drawn at random with a bias of 3/(3 + 2) in favor of the word (i.e., the probability of selecting the word was .6, and the probability of selecting the part-word was .4). If a word or a part-word were not in the internal lexicon of the model, its weight was considered to be null, and the same procedure was applied. This means that if only the word or only the part-word was in the internal lexicon, it was always selected. If neither the word nor the part-word was in the lexicon, the response was drawn at random without probabilistic bias.

The test performed by the participants included bisyllabic units, in order to avoid the strategic search for trisyllabic units after the initial test. Dealing with the test pairs of bisyllabic items in the same way as the test pair of trisyllabic items (i.e., as a function of their presence in PARSER's lexicon) would have been inappropriate. Indeed, bisysllabic chunks may be created when learning is in progress, but they normally disappear by decay and interference when learning progresses (see Giroux & Rey, 2009, for empirical evidence). However, it is likely that human participants are able to recognize a short segment even if it does not match entirely a given lexical entry, and they may use this knowledge to address the task demand. For instance, if readers were asked to judge whether ''xical'' recently appeared in this study, it is likely that they would say ''yes'' on the basis of their memory of having read ''lexical'' a few lines ago. This would happen even though ''xical'' is presumably not internally represented as a unit. The same rationale was applied to PARSER. A bisyllabic unit was considered as learned not only when it was represented as such in the model's lexicon but also when it was a component of a learned trisyllabic unit. Because performance on bisyllabic items depends, at least partially, on the knowledge of trisyllabic items, results were pooled over bisyllabic and trisyllabic test items (note that there was no difference in human participants as a function of items).

3.3. Results

The selection of free parameters is an ubiquitous problem in computational research. The general strategy adopted in earlier studies using PARSER (e.g., Frank, Goldwater, Mansinghka, Griffiths, & Tenenbaum, 2007; Giroux & Rey, 2009; Perruchet & Peereman, 2004; Perruchet et al., 2004) has been to first apply the parameters used in the initial study (Perruchet & Vinter, 1998; hereafter those parameters will be referred to as the standard parameters). In many cases, no subsequent adjustment of the parameters is needed. However, there are cases in which using the standard parameters leads to observing ceiling or floor effects. This outcome can be fixed through a modification of the rate of forgetting. Because PARSER proceeds through the selection of candidate units, forgetting must be neither too fast (PARSER would forget everything) nor too slow (PARSER would retain everything). This strategy (i.e., initial use of standard parameters, eventually followed by an explicitly motivated parameters adjustment) does not offer a guarantee that the reported results are invariant across a broad range of parameter settings (Boucher & Dienes, 2003). However, this ensures at least that performance fitting is not the end result of trial-and-error attempts to improve fitting through arbitrary parameter adjustments (note that when PARSER's other parameters are changed for exploratory purposes, this usually does not alter the general pattern of results; e.g., Frank et al., 2007).

The results obtained with the standard parameters are shown in Fig. 2. Half of the scores were at or near to 100%, hence indicating ceiling effects. Perruchet and Vinter (1998) also noted that PARSER tends to outperform human performance when trained with the same material. We performed an ANOVA as for human participants with Group (IWL+ vs. IWL−)

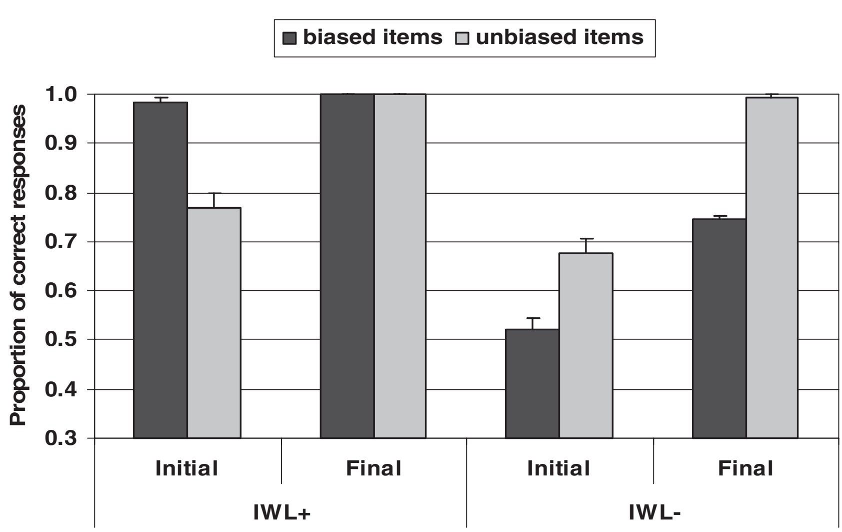


Fig. 2. Simulations from PARSER using the parameters of Perruchet and Vinter (1998). As for human participants, the figure reports the proportion of correct responses in the initial and the final tests, as a function of Groups, for biased and unbiased items. Error bars represent standard errors.

as a between-subjects factor and Test (initial vs. final) and Word status (biased vs. unbiased) as within-subjects factors. Although many effects were trivial or not interpretable due to ceiling effects, this analysis confirmed that (a) performance in the Group IWL+ was better than in the Group IWL−, $F(1, 38) = 269.28$, $p < .001$, $\eta_p^2 = 0.876$, hence indicating that the implementation of the perceptual biases was effective; and (b) performance improved between the initial test and the final test, $F(1, 38) = 230.90$, $p < .001$, $\eta_p^2 = 0.859$, hence attesting for learning. Learning also occurred when the analysis was restricted to the Group IWL− $F(1, 38) = 181.46$, $p < .001$, $\eta_p^2 = 0.905$, despite the fact that part-words were positively biased.

During the initial test, performance on the biased words was almost at maximum in the Group IWL+ and nearly at chance level for the Group IWL−. Crucially, this effect transferred to the unbiased words, despite the fact that these words were identical between groups. The effect of bias on the unbiased words was attenuated, but it was still significant, $F(1, 38) = 4.66$, $p = .037$, $\eta_p^2 = 0.109$. Thus, the effect of IWL was certainly excessive in the simulation, but the same trends as in human participants were observed.

To decrease the efficiency of learning in PARSER with the aim of removing ceiling effects, the rate of forgetting must be increased. This may be performed by manipulating either the rate of decay, the rate of interference, or both of them. For the sake of simplicity, only the rate of decay was manipulated in the present study. In addition, the action of IWL was made probabilistic, in order to reduce the influence of this variable on performance. We ran a set of simulations in which (a) the rate of decay was gradually increased (with a step of .005) starting from the standard value (.05), until the disappearance of ceiling effects and (b) the selection of a positively biased unit (among possible choices at a given processing step) was constrained on a probabilistic basis. The probability was set to .5, which means that on a given processing step, there was 50% chance for the selection of the positively biased unit to be constrained (as explained above). When the selection was not constrained, the usual algorithm was applied (i.e., the selection of the positively biased unit was neither compelled nor prevented).

Fig. 3 reports the results collected with a forgetting rate of .085, which was the first value for which performance was inferior to 100% in all conditions. An ANOVA was performed with Group (IWL+ vs. IWL−) as a between-subjects factor and Tests (initial vs. final) and Word status (biased vs. unbiased) as within-subjects factors. As expected, the Group IWL+ outperformed the Group IWL−, $F(1, 38) = 104.52$, $p < .001$, $\eta_p^2 = 0.733$, and final performance was better than initial performance, $F(1, 38) = 93.88$, $p < .001$, $\eta_p^2 = 0.712$. The interaction between the two factors was significant, $F(1, 38) = 31.47$, $p < .001$, $\eta_p^2 = 0.453$, indicating that language exposure was more beneficial to the Group IWL+ than to the Group IWL−. However, planned analysis showed that learning remained significant when the analysis was restricted to the Group IWL−, $F(1, 19) = 9.00$, $p = .007$, $\eta_p^2 = 0.321$.

As for human participants, the Group × Test interaction did not differ as a function of whether the biased items or the unbiased items were considered, as indicated by the nonsignificant Group × Test × Word Status interaction, $F(1, 38) = 0.51$, $p = .479$, $\eta_p^2 = 0.013$. Analyses performed separately on biased items and on unbiased items showed that language exposure was more beneficial to the Group IWL+ than to the Group IWL− for each of the

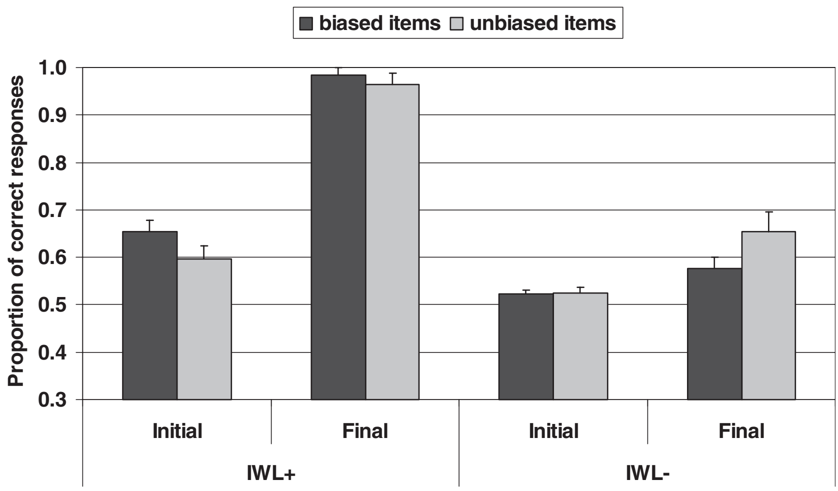


Fig. 3. Simulations from PARSER after the forgetting rate has been increased until the performance was inferior to 100% in all conditions, and with the effect of IWL having been made probabilistic. As in Fig. 2, the figure reports the proportion of correct responses in the initial and the final tests, as a function of Groups, for biased and unbiased items. Error bars represent standard errors.

two categories of items—biased items: $F(1, 38) = 88.16$, $p < .001$, $\eta_p^2 = 0.699$; unbiased items: $F(1, 38) = 65.52$, $p < .001$, $\eta_p^2 = 0.633$. Also as for human participants, the interaction between Group and Word Status was significant, $F(1, 38) = 12.16$, $p < .001$, $\eta_p^2 = 0.242$. For the Group IWL+, correct responses were more numerous for the biased items than for the unbiased items, $F(1, 19) = 5.19$, $p = .034$, $\eta_p^2 = 0.215$, whereas for the Group IWL−, correct responses were more numerous for the unbiased items, $F(1, 19) = 7.32$, $p = .014$, $\eta_p^2 = 0.278$.

An overall comparison between Fig. 1 (human data) and Fig. 3 (simulation) indicates that PARSER was successful in generating the pattern observed in human participants without implementing any ad-hoc algorithmic changes. Learning for biased items was better for the Group IWL+ than for the Group IWL−, even though our measure of change (i.e., the comparison between the two tests) controlled for the direct effect of IWL on word/part-word selection, and this difference transferred to the unbiased items.

4. Discussion

4.1. The interplay of statistical cues, IWL, and contextual information

In the reported experiment, participants were exposed to an artificial language in which word segmentation could be influenced by three sources of information. The first one was statistical information, and more specially the fact that word internal syllables are more cohesive than syllables spanning word boundaries, as in all studies patterned after the design of Saffran and collaborators (e.g., Saffran et al., 1996). This feature was identical for all participants. The second source of information was the relative IWL of words and part-words. For three of the six words, the selection was made from the results of a prior study (Perruchet et al., 2004) in such a way that, for half of the participants (Group IWL+), the words were spontaneously perceived as perceptual units in a continuous speech flow more often than the part-words generated by the succession of the words, whereas for the other half of the participants (Group IWL−), part-words were spontaneously perceived as perceptual units more often than the words. The third source of information concerned the other three words, which were perceptually unbiased. For the Group IWL+, information to extract these unbiased words was provided by the positively biased words, which were assumed to be identified earlier during the familiarization phase than the unbiased words. The Group IWL−, for which the corresponding information was not available, served here as a control.

The main result was that, after language exposure, the increase in performance in a forced choice segmentation task was larger for the Group IWL+ than for the Group IWL−, indicating that word discovery was affected by the IWL. The exploitation of statistical regularities was more efficient when IWLs fit with the language structure than when they contradicted this structure. Crucially, this facilitation transferred to the unbiased words, indicating that word discovery was also affected by the context. Indeed, learning of unbiased words was better when these words were mixed with positively biased words (Group IWL+) than when these words were mixed with negatively biased words (Group IWL−), despite the fact that these unbiased words were the same for the two groups of participants, and hence shared all of their statistical and perceptual properties.

Although the joint influences of the three sources of information considered here had not been previously tested, the effects of these factors taken individually have received a different amount of experimental support in earlier studies. The exploitation of statistical regularities in word segmentation has been well-documented since the studies by Saffran and collaborators (e.g., Saffran et al., 1996). The influence of phonological and prosodic cues on this type of learning has also been demonstrated (e.g., Creel et al., 2006; Onnis et al., 2005; Shukla et al., 2007; Tyler et al., 2006), but with a different experimental method. In earlier studies, some acoustical cues were added to the otherwise perceptually flat language during the familiarization phase. They were withdrawn from the items played during the test, to prevent these cues from *directly* influencing the responses during the test. In our experiment, the very same stimuli were played during the familiarization phase and during the test phase. The direct influence of perceptual cues on the forced-choice test was controlled by comparing performance collected after a short exposure to performance after longer exposure, on the assumption that the direct influences of these cues were identical in the two tests. Finally, the positive effect of knowing some words of the language in discovering the other words has also received some earlier support from the studies by Bortfeld et al. (2005) and Dahan and Brent (1999), although in very different conditions. In those earlier studies, participants were assumed to have acquired a given word (either a familiar word such as *Mommy* for the infants in Bortfeld et al.'s study, or a nonsense word presented in isolation in the study of Dahan and Brent), and the training utterances consisted of short sequences comprising this word in combination with a single other new word. In this paradigm, everything has been arranged to facilitate the exploitation of the known word by the learner. Our present experiment involved a more complex experimental setting. All the words were mixed within a single continuous sequence, and they only differed with regard to their relative IWL. The influence of the earlier extracted words on the discovery of the other words suggests that the importance of this process in word segmentation is somewhat undervalued in the current literature (but see Brent, 1999).

Our hypothesis was that PARSER (Perruchet & Vinter, 1998) could account for the joint influences of the three sources of information considered above. Indeed, the model relies on the formation of attentional chunks, the content of which partly depends on IWL. Moreover, because internal representations guide the partitioning of the sensory input, the early formation of some chunks naturally guide the model toward the discovery of other chunks. These predictions were tested by entering the language heard by the participants into PARSER. The only change to the original model (Perruchet & Vinter, 1998) was that the initial selection of the candidate units was biased by IWL. Positively biased units (whether they were words or part-words) were preferred to other candidate units while respecting the other constraints of the model. When the standard parameters used in Perruchet and Vinter (1998) were applied, ceiling effects were observed. When forgetting was progressively increased until ceiling effects disappeared in all conditions, the pattern of results obtained by PARSER reproduced the main effects observed for human participants.[3]

Below we examine whether other models of word segmentation can simulate these results. We then discuss the implications of our study for natural language processing, by considering word segmentation and more general aspects of early language acquisition.

4.2. The computation of transitional probabilities

Certainly, the prevalent view on word segmentation is based on the suggestion by Saffran et al. (1996; see also Aslin et al., 1998) that boundaries between words can be found through the exploitation of transitional probability. Participants would exploit the fact that, on average, the transitional probabilities between word internal syllables are stronger than the transitional probabilities between syllables spanning word boundaries. This view fits well with the widespread use of Simple Recurrent Networks (SRN, e.g., Christiansen, Allen, & Seidenberg, 1998; Elman, 1990;) to simulate segmentation processes. The use of more sophisticated measures of statistical association has been suggested too, in particular by Swingley (2005).

Most researchers investigating the role of statistics fully acknowledge that word segmentation may be influenced by other, independent factors (e.g., Aslin et al., 1998; Gomez, 2007; Seidenberg & MacDonald, 1999; Thiessen & Saffran, 2003). The crucial point, however, is that the way statistical computations combine with other cues is not constrained by these models and can be construed ad lib in quite divergent ways.

Concerning the effects of contextual factors, the role of knowing some words of the language in the discovery of others has never been integrated in word segmentation approaches involving statistical computation. A transitional probability, for instance, is a value inherent to a pair of syllables, which does not depend on whether the local context in which this pair of syllables appears is known by the learner. As asserted by Dahan and Brent (1999), ''transitional-probability computations do not take into account the segmentation points in previous utterances; in other words, having isolated some words does not help in isolating other words or even the same words later on'' (p. 166).

Concerning the influence of IWL, the situation is different. The relation between statistical structure and prosodic or phonological information—a part of what has been subsumed above under the term IWL—has been often envisioned, but, importantly, in very different ways. The action of different cues can be thought of as being mediated by independent processes, which would operate in parallel. Statistical computations would be blind to the perceptual properties of the material. This is the view advocated by Shukla et al. (2007), at least with regard to prosody. The authors suggest that transitional probability computations (or other forms of statistical computation) over syllabic representations of speech rely on encapsulated, automatic processes, which proceed irrespective of the prosodic break-points. Prosody would act subsequently as a filter, suppressing possible word-like sequences that straddle two prosodic constituents. Another possibility is that statistical learning is ''guided'' by the initial perceptual biases. It has long been claimed that the exploitation of statistical regularities needs to be constrained by external factors. The acoustical properties of the speech flow could serve as such constraints (e.g., Gomez, 2007; Onnis et al., 2005; Saffran, 2002; Seidenberg & MacDonald, 1999). Still another view is that statistical computations would be performed on representations that embed prosodic or phonological information. Curtin et al. (2005), for instance, suggest that stressed and unstressed syllables with the same segmental content could be considered as different primitives for the computation of transitional probabilities.

To sum up, approaches based on the assumption that learners perform some kind of statistical computation generate no specific prediction regarding the effects of contextual and perceptual cues.[4]

These effects can only be accounted for a posteriori, by postulating the existence of additional, independent processes. Two earlier studies, which compared PARSER and connectionist models with regard to their ability to account for other variables influencing learning, also lead to the conclusion that connectionist models would require specifically designed modifications in conditions where the standard version of PARSER simulates the behavioral data patterns. Perruchet and Desaulty (2008) showed that participants were able to learn the words from an artificial language when the only available cues were backward transitional probabilities. The authors showed that PARSER accounts for this ability, while an SRN is only sensitive to forward transitional probabilities. Likewise, Giroux and Rey (2009) tested the predictions of SRN and PARSER in a situation where the recognition of words and sublexical units were examined after 2 or 10 min of exposure to an artificial speech stream. Both models predict similar performance on words and sublexical units after 2 min. However, only PARSER predicts better performance on words relative to sublexical units after 10 min. Performance of human participants were consistent with PARSER's predictions. In both studies, PARSER accounted for the data without any change to the original version of the model (Perruchet & Vinter, 1998), even though this version was obviously not designed to deal with these situations. It is likely that a connectionist model would be also able to simulate these data, but, in contrast with PARSER, this achievement would need ad-hoc algorithmic modifications.

4.3. Brent's INCDROP and the MDL-based models

In Brent's INCDROP model (Brent, 1996; Brent & Cartwright, 1996), as in PARSER, the segmentation of a new syllable sequence is dependent on the units stored in the lexicon. If there is no possible mapping between some part of the sequence and the stored units, the whole utterance is stored in memory as a single unit. Otherwise, the familiar units are exploited to extract new potential units from the utterance. The choice between different possible segmentations of an utterance is construed as an optimization problem. The principle of the method is akin to establishing a list of the possible segmentations of a given utterance (although the authors used computational tools that prevented the program from proceeding in this way). The choice between possible segmentations is then made in order to fulfill a number of criteria. These criteria are threefold (according to Brent, 1996): minimize the number of novel words, minimize the sum of the lengths of the novel words, and maximize the product of the relative frequencies of all the words. The process of optimization is performed thanks to a statistical inference method, called the MDL method. Other models of word segmentation also rely on MDL-based algorithms (e.g., de Marcken, 1996; see also the MDLChunker model by Robinet & Lemaire, 2009), often within a Bayesian framework (e.g., Brent, 1999; Goldwater, Griffiths, & Johnson, in press).

INCDROP bears several striking similarities with PARSER. In contrast to models based on the computation of transitional probabilities, these two models posit the primacy of chunks. MDL methods exhaustively examine all possible partitionings of the corpus, while PARSER relies on the variety generated by successive random drawings to provide provisional chunks, but this difference can hardly be thought of as a crucial one. In both models, the segmentation problem is solved by some direct competition between different possible chunks, instead of being an inference on a continuous distribution of probabilities over syllables. As PARSER, INCDROP has the capacity of accounting for the action of multiple cues. The optimization algorithm makes the model sensitive to the statistical structure of the speech, and the power of lexically driven segmentation is fully exploited. In addition, the influence of IWL can be easily implemented. For instance, in Brent and Cartwright (1996), the choice between possible segmentations takes account of certain phonotactic constraints in English words. Finally, it is highly likely that INCDROP would be able to account for the recent results outlined above, without major changes. For instance, there is no principled reason for an MDL-based model to be limited to the exploitation of forward transitional probabilities and hence, INCDROP would be certainly able to account for the results of Perruchet and Desaulty (2008). Likewise, Robinet and Lemaire (2009) show that their MDLChunker model, which relies on the same basic principles as INCDROP, is able to reproduce Giroux and Rey (2009)'s effect, which was initially taken as a support for PARSER.

To sum up, although further studies would be necessary to confirm this assertion, it seems rather difficult to separate PARSER from the MDL-based models with regard to their explanatory power (as revealed in current empirical studies). If there is no easy way to distinguish between models on empirical grounds, on which alternative bases may rely a preference judgment? A common test-bed for choosing between theories that are equally consistent with the data relies on the principle of simplicity or parsimony (e.g., Chater & Vitanyi, 2003). However, this principle has been claimed to be at the core of the motivation underpinning both PARSER and MDL-based models. PARSER instantiates the parsimony principle at the level of the psychological processes engaged by the learner to cope with a given material. To meet this objective, it avoids introducing ad-hoc mechanisms and processes and construing the properties of the description of the data as a driving force. In contrast, MDL-based models instantiate the parsimony principle at the level of the description of the data. They privilege the more economical description of the data patterns, without considering the complexity of the algorithms the learners have to use to discover this mode of representation.

Rather than speculating on the more relevant implementation of parsimony—which can be seen as a matter of opinion—we would like to briefly discuss the following observation: Applying the principle of parsimony to the learner's abilities or to the learner's final representation of the data leads to the same outcome, at least for word segmentation. Worthy to note, this is a logical consequence of the models' organization. In a nutshell, MDL-based models select the mode of segmentation that generates the minimum number of different words. The crucial point is that for a fixed corpus, aiming for a minimum number of different words maximizes the number of repetitions of each word. PARSER exploits this logical corollary, relying on the fact that selecting the more frequent words among a set of candidate units is the natural by-product of ubiquitous laws of forgetting. Extracting the more frequent units of a corpus naturally leads PARSER to describe this corpus with the minimum number of different words.

4.4. Implications for understanding word segmentation

Investigating and modeling word segmentation in artificial languages are obviously aimed at improving our understanding of natural language acquisition. Regarding IWL, one may wonder about the adaptive value of the effects evidenced in our study. In comparison to the segmentation of unbiased words, the segmentation of positively biased words is facilitated, but the segmentation of negatively biased words (i.e., words for which their combination generates positively biased part-words) is impeded roughly to the same extent, if not more (as shown in Fig. 1). This might suggest that these opposite trends cancel each other out. However, this would be true only if there were as many negatively biased words as positively biased words in natural languages. This is quite unlikely. As far as first language acquisition is considered, phonological and prosodic features generally provide a reliable starting point, because most of them are themselves learned through language exposure. Thus, their positive consequences should, on average, largely exceed their negative ones. Note that the positive interplay of statistical and perceptual factors for segmenting the words of natural languages has been emphasized earlier (e.g., Swingley, 2005).

Our result showing the exploitation of known words for discovering new words provides a further innovative contribution. Although unbiased items were identical in the two groups of participants, they were learned better when they were mixed with positively biased items than when they were mixed with negatively biased items. The exploitation of already-known words might be considered as a somewhat peripheral effect, just being able to account for some occasional fine-tuning. We argue that this view is faulty, for at least two reasons. First, being exposed to a language in which all words are new (as in laboratory conditions) is certainly quite exceptional in real-world settings. As a consequence, if learners are actually able to benefit from known words, this ability could be exploited in an overwhelming proportion of utterances. Second, exploiting already known words could be more than an optional complement to other mechanisms, but it could be rather essential for the full achievement of word discovery. Indeed, neither statistical nor perceptual factors can be construed as decisive when considered in isolation. The joint consideration of the two categories of cues certainly improves performance considerably (e.g., Onnis et al., 2005), but it does not lead to perfect performance. This is a logical consequence of the fact that both categories of cues are based on probabilistic information. What about the words that fit neither with statistics nor with the dominant prosodic or phonological patterns of the language? Our response is that they may be easily learned, because learners use the information provided by the surrounding words (i.e., the contextual information).

Our final claim is that to be fully understood, the investigation of word segmentation needs to consider the dynamical interplay between (at least) the three sources of information we have investigated in this study. This claim could suggest that a full theoretical account of word segmentation would require a number of unrelated mechanisms, each of them devoted to exploit a specific part of the available information. Our modeling approach provides a more parsimonious solution. A simple, unified model can account for the dynamic interplay between the statistical structure of the ongoing input, the IWL of parts of the speech flow, and the contextual information provided by the earlier emergence of other word-like unit. Note that we are not arguing that this model, PARSER, such as currently implemented for the processing of miniature languages composed of a small set of syllables, would be able to extract all words from a natural language. PARSER was not aimed at providing a realistic, full-scale model of word segmentation. Our objective was rather to show that the basic principles of associative learning and memory (on which PARSER relies) are able to account for the overall dynamics of word extraction, without introducing any ad-hoc mechanisms or specialized modules.

4.5. Toward a generalization to early language acquisition

Word extraction from a syllabified input is itself just one component of early language acquisition, and the question arises of whether the conclusions reached for this particular step of processing, and notably the explanatory power of basic principles of associative learning and memory, can be generalized. Although any response to this question would be premature at this stage, we wish to provide fuel for a positive response in the final part of this study by examining the ability of PARSER to account for other parts of language acquisition. Indeed, the field of relevance of the PARSER model is not limited to the specific stage of language acquisition that was explored here.

Let us consider first the word-likeness of certain sequences of syllables, which was entered into the model as preestablished knowledge. To some extent, IWL can be construed as the end-result of earlier applications of PARSER's principles. We have pointed out in the introduction that the IWL of a given set of sounds could be linked to its similarity with one or several words from the learners' native language. As PARSER is primarily devised to account for word formation, it is possible to conceive the discovery of these words as the end product of PARSER's principles. More certainly, IWL is also determined by phonological and prosodic features, such as lexical stress placement (e.g., Creel et al., 2006; Curtin et al., 2005; Thiessen & Saffran, 2007), which are, at least in part, dependent on the learners' experience with their mother tongue. PARSER's principles could be also relevant here, when this perceptual information is integrated into the primitives that are given to the model.

Perruchet and Peereman (2004) have shown experimental and computational evidence for this hypothesis. The authors focused on the relationship between vowel (V) and consonant (C) in the terminal syllables of words as a determinant of their word likeness. Focusing on the phonemes composing the rimes (VC) is justified by the fact that analyses of linguistic corpora suggest that there are strong probabilistic constraints on the VC combinations of rimes, at least in English (Kessler & Treiman, 1997) and in French (Peereman, Dubois-Dunilac, Perruchet, & Content, 2004). Perruchet and Peereman's experiment revealed that VC terminal endings were a reliable determinant of how well nonwords sounded like French words, for both children and adults. More importantly for the present concern, the authors examined the ability of computational models to account for these results, successively considering a connectionist model (SRN; e.g., Elman, 1990) and PARSER. PARSER was a better predictor of performance than the SRN. In particular, PARSER was sensitive to the same measures of statistical consistency that human participants exploited to produce their word-likeness judgments. Of course, we are not claiming that all determinants of IWL are shaped by PARSER-like processes, and further research is required to examine the extent to which Perruchet and Peereman's conclusions may be generalized, but there is no a priori reason to anticipate that rimes would be a particular case.

Let us examine now how the principles on which PARSER is grounded may also be relevant for some steps of processing that follow word extraction. In the line of Saffran et al. (1996) studies, our present experiment investigated how listeners create units composed of sound sequences. In natural settings, a further step of processing consists in mapping the sound sequences to their referents. It may be argued that the relation between a word and its referent is a form of association and, hence, that a model able to learn associations, as PARSER is, is well suited to achieve such a mapping. This associative view has been criticized on the ground of the indeterminacy of the referent in concrete situations, and various specific constraints have been proposed to solve the indeterminacy problem. Perruchet and Vinter (2002, Section 4.2; see also Perruchet, 2005) have suggested how PARSER's logic, when extended to the word-referent mapping, could offer a solution without postulating multiple constraints. In the same line, Smith and Yu (2008; see also Yu & Smith, 2007) have recently provided experimental evidence supporting the view that statistical learning could underpin word-referent mappings. They showed that infants are able to exploit the statistical information embedded in a set of individually ambiguous scenes to map words to their referents. Interestingly, they refer to our approach of word segmentation when they note that

> statistical learning need not be the result of highly specialized statistical learning mechanisms (e.g., Perruchet & Vinter, 1998). The learner could solve this learning task via simple (…) associative learning mechanisms. Across trials, the learner could accumulate associations between words and potential referents by strengthening and weakening associative links with each co-occurrence or non-co-occurrence. (Smith & Yu, 2008, p. 1566)

Looking one step further, PARSER is not necessarily limited to the word level. It has been successfully applied, with only minor parametric adjustments, to multiword units (e.g., the transcoding of numbers from their verbal to their digital forms, Barrouillet et al., 2004) and to the learning of finite-state grammars (Perruchet et al., 2002; Pothos, 2007). This brief outline (see Perruchet & Vinter, 2002; for a more developed account) is sufficient to make our point: PARSER is not limited to the formation of word-like units from syllabic primitives. Its area of relevance goes from the subsyllabic level (Perruchet & Peereman, 2004) to rudiments of syntax (e.g., Perruchet et al., 2002), and it may be easily generalized to the fundamental issue of word-referent mappings (Smith & Yu, 2008). By illustrating this property of PARSER, we do not intend to pinpoint an advantage that would be specific to this model. Extensive domains of application and generality have been reported also for other models, such as connectionist models and MDL-based models (see for instance de Marcken, 1996, regarding some aspects of syntactic structure). Our intent was rather to provide a significant support to the idea that general, all-purpose processes such as those implemented in PARSER are able to account for a larger part of early language acquisition than once thought.

Notes

1. It could be argued that the model is a priori irrelevant for word segmentation, because it is based on the attentional processing of the ongoing information while word segmentation would be an automatic or implicit form of learning. Many studies have investigated the role of attention in statistical/implicit learning, and there is growing consensus that this form of learning requires attention (for a review, see Perruchet, 2008). There is no principled reason to think that segmentation studies in the line of Saffran et al. would differ from other implicit learning paradigms in this respect. Observing learning under dual-tasks conditions (as for instance in Saffran et al., 1997) does not imply the existence of a nonattentional form of learning, because the secondary task may not deplete the attentional resources completely. Moreover, the conclusion according to which the discovery of relevant units requires attention as any other

forms of learning has been supported by studies using continuous speech flow (Toro et al., 2005) and visual shapes (e.g., Baker et al., 2004; Turk-Browne et al., 2005).

2. Decay and interference are two different ways for simulating forgetting. Interference refers to *specific* (retroactive) interference, that is, to the detrimental effect of processing items similar in some ways to stored memories. A decay parameter was introduced to simulate the fact that the memory for any item tends to decrease over time, even if there is no specific interfering event in the subsequent input. However, the model makes no claim about the existence of a genuine decay process. Decay may be thought of alternatively as *nonspecific* interference.
3. Given that IWL was hand-coded in the model, its effect on word discovery could be thought of as self-evident. However, if attributing positive IWL to a set of syllables would have led to endorse this set of syllables as a word through a direct link of causality, the part-words would have been mistakenly identified as words when PARSER processed the language played to Group IWL−, given that for this group, the positively biased units were the part-words. In contrast with this prediction, PARSER (as human participants) was successful in discovering the words.
4. As an aside, it is worth adding that PARSER exploits the statistical structure of the material as well as specially designed computational models. Hunt and Aslin (2001) wondered about the ability of PARSER to discover words in a frequency-balanced task. They pointed out that Perruchet and Vinter (1998) did not model the experiment by Aslin et al. (1998), in which infants learned the word of a language in which frequency was controlled, presumably exploiting transitional probabilities between syllables. Although it is correct that Perruchet and Vinter did not address the issue, it would be incorrect to claim that PARSER is unable to account for Aslin et al.'s results. In fact, PARSER, without any ad-hoc modification of the original algorithm, is sensitive to more sophisticated measures of association than raw frequency, including transitional probability, and moreover bidirectional measures of association. This somewhat surprising ability is due to the fact that forgetting is simulated by both decay and interference, as spelled out elsewhere (Perruchet & Desaulty, 2008; Perruchet & Pacton, 2006; Perruchet & Peereman, 2004).

Acknowledgments

This work has been supported by the Centre National de la Recherche Scientifique (CNRS, UMR 5020 and 5022), as well as by the Université de Bourgogne and the Université Claude Bernard Lyon 1.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). *Role of attention and perceptual grouping in visual statistical learning*. *Psychological Science*, *15*, 460–466.

Barrouillet, P., Camos, V., Perruchet, P., & Seron, X. (2004). *ADAPT: A Developmental, Asemantic, and Procedural model for Transcoding from Verbal to Arabic numerals*. *Psychological Review*, *111*, 368–394.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). *Mommy* and me. *Psychological Science*, *16*, 298–304.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*, 807–842.

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition*, *61*, 1–38.

Brent, M. R. (1999). peech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, *8*, 294–301.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Case, R. (1993). Theories of learning and theories of development. *Educational Psychologist*, *28*, 219–233.

Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.

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

Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 15–32.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, *96*, 233–262.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, *128*, 165–185.

Dutoit, T., Pagel, N., Pierret, F., Bataille, O., & Van Der Vrecken, O. (1996). The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings of the ICSLP'96* (Vol. 3, pp. 1393–1396), Philadelphia.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fischer, K., & Granott, N. (1995). Beyond one-dimensional change: Parallel, concurrent, socially distributed processes in learning and development. *Human Development*, *38*, 302–314.

Frank, M. C., Goldwater, S., Mansinghka, V., Griffiths, T., & Tenenbaum, J. (2007). Modeling human performance on statistical word segmentation tasks. *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 281–286), Austin, TX: Cognitive Science Society.

Giroux, I., & Rey, A. (2009). Lexical and sub-lexical units in speech perception. *Cognitive Science*, *33*, 260–272.

Goldwater, S., Griffiths, T. L., & Johnson, M. (in press). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*.

Gomez, R. (2007). Statistical learning in infant language development. In M. Gareth Gaskel (Ed.), *The Oxford handbook of psycholinguistics* (pp. 601–615). New York: Oxford University Press.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*, 658–680.

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

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: Bradford/MIT press.

Kessler, B., & Treiman, R. (1997). Syllable structure and phoneme distribution. *Journal of Memory and Language*, *37*, 295–311.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376–400.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicon. *Journal of Experimental Psychology: General*, *132*, 202–227.

de Marcken, C. G. (1996). *Unsupervised language acquisition*. PhD thesis, MIT (available at: http://www.demarcken.org/carl/papers/PhD.pdf) accessed October 1, 2009.

Matthys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477–500.

Onnis, L., Monaghan, P., Chater, N., & Richmond, K. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*, 225–237.

Peereman, R., Dubois-Dunilac, N., Perruchet, P., & Content, A. (2004). Distributional properties of language and sub-syllabic processing units. In P. Bonin (Ed.), *Mental lexicon* (pp. 215–235). New York: NovaScience Publishers.

Perruchet, P. (2005). Statistical approaches to language acquisition and the self-organizing consciousness: A reversal of perspective. *Psychological Research*, *69*, 316–329.

Perruchet, P. (2008). Implicit learning. In H. L. Roediger, III (Ed.), *Coginitive psycology of memory: Vol. 2, Learning and memory: A Comprehensive reference*, 4. vols. (pp. 597–621). Oxford, England : Elsevier.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*, 1299–1305.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: Two approaches, one phenomenon. *Trends in Cognitive Sciences*, *10*, 233–238.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, *17*, 97–119.

Perruchet, P., Tyler, M. T., Galland, N., & Peereman, R. (2004). Learning non-adjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, *133*, 573–583.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness (target paper). *Behavioral and Brain Sciences*, *25*, 297–388.

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, *55A*, 485–503.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*, 227–244.

Robinet, V., & Lemaire, B. (2009). MDLChunker: A MDL-based model of word segmentation. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual meeting of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172–196.

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

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). *Incidental language learning*. *Psychological Science*, *8*, 101–105.

Schyns, P. G., Golstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–53.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*, 553–554.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*, 1–32.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

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

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

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

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25–34.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552–556.

Tyler, M. D., Perruchet, P., & Cutler, A. (2006). A cross-language comparison of the use of stress in word segmentation. Poster presented at the 152nd meeting of the Acoustical Society of America, Honolulu, November 29.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.

Appendix A

Appendix Table 1 shows the items used during the forced-choice test for the participants of Groups IWL+ and IWL−. Italicized items are the words heard during the familiarization phase.

Regarding biased items, the trisyllabic words for the Group IWL+ were identical to the trisyllabic part-words for the Group IWL−, and they were the same for all participants of each group. The part-words for the Group IWL+ and the words for the Group IWL− differed as a function of whether the number of the participant was even or odd. The difference was related to the generation of part-words (given the words ''ABC'' and ''DEF'', part-words can be either ''BCD'' or ''CDE''). The bisyllabic items used during the test also differed as a function of the parity of the number of the participant (given the word ''ABC,'' bisyllabic subcomponents can be either ''AB'' of ''BC'').

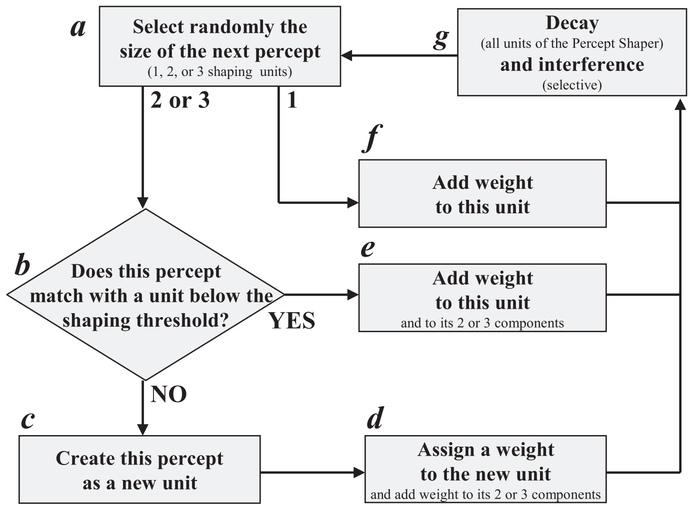
Regarding unbiased items, they were the same for two coupled participants from Group IWL+ and Group IWL−. However, the syllables were randomly permutated for each new pair of participants. The table presents the specific words and part-words heard by Participant 1 (#odd) and Participant 2 (#even) of each group. The differences again pertained to the mode of partitioning of the words.

Appendix B

PARSER is centered on a single vector, known as Percept Shaper (PS). PS is composed of the internal representations of the displayed material and may be thought of as a memory

Appendix Table 1

Item	Group IWL+				Group IWL-			
	N# Even		N# Odd		N# Even		N# Odd	
Status	Word	Part-word	Word	Part-word	Word	Part-word	Word	Part-word
Biased	*befoki*	kipuli	*befoki*	fokipu	*kipuli*	pulidu	*fokipu*	pulidu
	befoki	dutara	*befoki*	liduta	*kipuli*	taraga	*fokipu*	taraga
	befoki	gabefo	*befoki*	ragabe	*kipuli*	befoki	*fokipu*	befoki
	pulidu	kipuli	*pulidu*	fokipu	*dutara*	pulidu	*liduta*	pulidu
	pulidu	dutara	*pulidu*	liduta	*dutara*	taraga	*liduta*	taraga
	pulidu	gabefo	*pulidu*	ragabe	*dutara*	befoki	*liduta*	befoki
	taraga	kipuli	*taraga*	fokipu	*gabefo*	pulidu	*ragabe*	pulidu
	taraga	dutara	*taraga*	liduta	*gabefo*	taraga	*ragabe*	taraga
	taraga	gabefo	*taraga*	ragabe	*gabefo*	befoki	*ragabe*	befoki
	foki	kipu	befo	kipu	kipu	lidu	kipu	puli
	foki	duta	befo	duta	kipu	raga	kipu	tara
	foki	gabe	befo	gabe	kipu	foki	kipu	befo
	lidu	kipu	puli	kipu	duta	lidu	duta	puli
	lidu	duta	puli	duta	duta	raga	duta	tara
	lidu	gabe	puli	gabe	duta	foki	duta	befo
	raga	kipu	tara	kipu	gabe	lidu	gabe	puli
	raga	duta	tara	duta	gabe	raga	gabe	tara
	raga	gabe	tara	gabe	gabe	foki	gabe	befo
Unbiased	*ʒɛvydɛ̃*	dɛ̃sone	*ʒɛlõdɛ̃*	lõdɛ̃vy	*ʒɛvydɛ̃*	dɛ̃sone	*ʒɛlõdɛ̃*	lõdɛ̃vy
	ʒɛvydɛ̃	mãkøti	*ʒɛlõdɛ̃*	sotikø	*ʒɛvydɛ̃*	mãkøti	*ʒɛlõdɛ̃*	sotikø
	ʒɛvydɛ̃	lõʒɛvy	*ʒɛlõdɛ̃*	mãneʒɛ	*ʒɛvydɛ̃*	lõʒɛvy	*ʒɛlõdɛ̃*	mãneʒɛ
	sonemã	dɛ̃sone	*vysoti*	lõdɛ̃vy	*sonemã*	dɛ̃sone	*vysoti*	lõdɛ̃vy
	sonemã	mãkøti	*vysoti*	sotikø	*sonemã*	mãkøti	*vysoti*	sotikø
	sonemã	lõʒɛvy	*vysoti*	mãneʒɛ	*sonemã*	lõʒɛvy	*vysoti*	mãneʒɛ
	køtilõ	dɛ̃sone	*kømãne*	lõdɛ̃vy	*køtilõ*	dɛ̃sone	*kømãne*	lõdɛ̃vy
	køtilõ	mãkøti	*kømãne*	sotikø	*køtilõ*	mãkøti	*kømãne*	sotikø
	køtilõ	lõʒɛvy	*kømãne*	mãneʒɛ	*køtilõ*	lõʒɛvy	*kømãne*	mãneʒɛ
	ʒɛvy	dɛ̃so	ʒɛlõ	dɛ̃vy	ʒɛvy	dɛ̃so	ʒɛlõ	dɛ̃vy
	ʒɛvy	mãkø	ʒɛlõ	tikø	ʒɛvy	mãkø	ʒɛlõ	tikø
	ʒɛvy	lõʒɛ	ʒɛlõ	neʒɛ	ʒɛvy	lõʒɛ	ʒɛlõ	neʒɛ
	sone	dɛ̃so	vyso	dɛ̃vy	sone	dɛ̃so	vyso	dɛ̃vy
	sone	mãkø	vyso	tikø	sone	mãkø	vyso	tikø
	sone	lõʒɛ	vyso	neʒɛ	sone	lõʒɛ	vyso	neʒɛ
	køti	dɛ̃so	kømã	dɛ̃vy	køti	dɛ̃so	kømã	dɛ̃vy
	køti	mãkø	kømã	tikø	køti	mãkø	kømã	tikø
	køti	lõʒɛ	kømã	neʒɛ	køti	lõʒɛ	kømã	neʒɛ



Appendix Fig. 1. Operations performed by PARSER at each time step.

store or a mental lexicon. A weight, which reflects the person's familiarity with the item, is assigned to each element in PS. At the start of the familiarization session, PS contains only the primitives needed for the processing of the material (here, the syllables; for the reasons to consider syllables as processing primitives, see, e.g., Swingley, 2005). At the end, it should contain, in addition, the structurally relevant units that form the material (here, the words).

The way the words are built in PS during training is described in the flowchart in Appendix Fig. 1. Let us consider how the flowchart works for a language composed of six trisyllabic words, *befoki*, *pulidu*, *taraga*, *maneso*, *vujedin,* and *tilonke*. Let us assume that the sequence begins with *taragavujedinbefokitaragapulidu*…. The string is first segmented into small and disjunctive parts. In PARSER, the multiple determinants of this initial parsing are simulated by a random generator, which selects the size of the next percept within a range of 1–3 units (Appendix Fig. 1, step a). Suppose the random generator provides 2, 3, 2, 3, and 1 in the first five trials. As a consequence, the first percepts would be *tara*, *gavuje*, *dinbe*, *fokita*, and *ra*. Because none of the first four percepts is present in PS (step b), they are created as new units (step c) and assigned a weight (step d). Also, the weights of the components, *ta*, *ra*, *ga*, and so on, are incremented. The fifth percept, *li*, matches a primitive and hence is already represented in PS. Its weight is also incremented (step f).

At each time step (a time step is defined by the processing of a new percept, that is, by one cycle in the flowchart in Appendix Fig. 1), the units forming PS are affected by forgetting and retroactive interference (Appendix Fig. 1, step g). Forgetting is simulated by decreasing all the units by a fixed value. Interference is simulated by decreasing the weights of the units in which any of the syllables involved in the currently processed unit are embedded. In the case described here, interference occurs for the first time while *fokita* is perceived. Indeed,

Appendix Table 2
Changes in constituents of Percept Shaper during one time step

			Processing Components (Add Weight, Interference)				
	Initial Weights	Creation	ki	tara	ga	Forgetting	Final Weights
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ta	3.50			−0.005		−0.05	3.445
ra	3.10			−0.005		−0.05	3.045
ki	2.55		+0.5			−0.05	3.00
be	2.50					−0.05	2.45
fo	2.50					−0.05	2.45
vu	2.45					−0.05	2.40
ga	2.40				+0.5	−0.05	2.85
je	1.70					−0.05	1.65
tara	1.58			+0.5		−0.05	2.03
du	1					−0.05	0.95
dufoki	1					−0.05	0.95
kitaraga	—	1					1

Note. In this example, the changes are due to the perception of *kitaraga*, when *ki*, *tara*, and *ga* are existing units of the system.

ta is already present in an old unit in PS, *tara.* In consequence, the weight of *tara* is decremented when *fokita* is perceived (in addition to the decrement due to forgetting).

In this early phase, perception is driven by the initial primitives of the system, namely the syllables. However, the psychological principles implemented by the model stipulate that a representation created during learning may become able to guide perception, as the initial primitives were. The condition for an element of PS to shape perception is that its weight is at least equal to a threshold value. In contrast, when the frequency of perceiving a given element is not high enough to counteract the effects of forgetting and interference, this element is removed from PS when its weight becomes null.

In the reported simulations, the starting weight given to any created unit was set to 1. The primitives that formed PS before training were also assigned a weight of 1. The increment received by an old unit in PS when this unit serves to shape perception was set to 0.5. The decrements due to forgetting and interference were set to 0.05 and 0.005, respectively, for the first simulation (forgetting was increased for subsequent simulations, see main text). The last parameter, namely the threshold above which a unit is able to shape perception, was set to 1.

Let us assume that after having processed a small part of the language, the content of PS and the weight of each unit are as shown in Appendix Table 2, columns 1 and 2, respectively (units with a weight lower than 1 are not reproduced here). Let us also assume that the continuation of the sequence is *kitaragapulidu*…, and that the random generator (Appendix Fig. 1, step a) determining the number of units embedded in the next percept provides a value of 3. Perception would be shaped by the three units *ki*, *tara*, and *ga.*

Because *kitaraga* does not match a unit that has a weight below the shaping threshold (step b), this percept is created as a unit in PS (step c) and assigned a weight of 1 (step d; see Appendix Table 2, column 3). In addition, the components forming *kitaraga* receive an additional weight of 0.5 (Appendix Table 2, columns 4, 5, and 6). Note that *tara* is incremented, in keeping with the fact that it is an autonomous component of the percept *kitaraga*, but not its primitive parts *ta* and *ra*, which do not share this status. On the contrary, *tara* interferes with *ta* and *ra* and, in consequence, the weights of these units are decremented by 0.005. Finally, all the units in PS are decremented by 0.05 to simulate forgetting (Appendix Fig. 1, step g; see Appendix Table 2, column 7). The rightmost column of Appendix Table 2 displays the state of PS after *kitaraga* has been processed.