J. Child Lang. **45** (2018), 204–218. © Cambridge University Press 2017
doi:10.1017/S0305000916000660

BRIEF RESEARCH REPORT

Reduplication facilitates early word segmentation*

MITSUHIKO OTA AND BARBORA SKARABELA
University of Edinburgh, UK

(*Received* 20 *July* 2016 – *Revised* 8 *November* 2016 – *Accepted* 12 *December* 2016 – *First published online* 6 *February* 2017)

ABSTRACT

This study explores the possibility that early word segmentation is aided by infants' tendency to segment words with repeated syllables ('reduplication'). Twenty-four nine-month-olds were familiarized with passages containing one novel reduplicated word and one novel non-reduplicated word. Their central fixation times in response to these as well as new reduplicated and non-reduplicated words introduced at test showed that familiarized reduplicated words were segmented better than familiarized non-reduplicated words. These results demonstrate that infants are predisposed to segment words with repeated phonological elements, and suggest that register-specific words in infant-directed speech may have evolved in response to this learning bias.

INTRODUCTION

Infants at the cusp of word learning must identify units of sounds that constitute words in the ambient running speech. This is not a trivial task, as most words in natural speech occur in multiword utterances where boundaries between words are not typically marked by a break in the signal (Aslin, Woodward, LaMendolda & Bever, 1996; Klatt, 1989). Research on this problem to date has uncovered a range of mechanisms that facilitate infants' segmentation of words from fluent speech input. For example, infants can deploy their ability to track transitional probabilities to group sound units that are more likely to constitute words (Saffran, Aslin & Newport, 1996). Infants may also learn the general phonological properties of the words in the language, such as their predominant stress

[*] This research was supported by the UK Economic and Social Research Council (ES/J023825/1). Address for correspondence: Mitsuhiko Ota, University of Edinburgh – Linguistics and English Language, Dugald Steward Building, Edinburgh EH8 9AD, United Kingdom. e-mail: mits@ling.ed.ac.uk

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

pattern and phonotactic constraints, and then utilize that information to identify word forms (Johnson & Jusczyk, 2001; Jusczyk, Hohne & Bauman, 1999; Jusczyk, Houston & Newsome, 1999; Mattys, Jusczyk, Luce & Morgan, 1999; Saffran & Thiessen, 2003).

Recent research has also demonstrated that word segmentation can be aided by words already known to the learner. For instance, infants are able to use a few previously acquired words such as *mommy/mama* and their own name to identify and segment adjoining words in running speech (Bortfeld, Morgan, Golinkoff & Rathburn, 2005; Sandoval & Gómez, 2016). This shows that a small set of words that are acquired early can bootstrap lexical development by acting as anchor points for further word segmentation.

But how are the anchor words learned in the first place? In the case of words such as *mommy/mama* and the infant's given name, it may be their extremely high input frequencies, combined with a general learning mechanism such as transitional probability tracking, that makes them early-acquired. In this paper, we examine a further, previously unexplored possibility that some words may be acquired early (and hence can anchor further segmentation) because their own phonological characteristics make them inherently more likely to be segmented in running speech. In particular, we test the hypothesis that segmentation is facilitated in words containing reduplication, or the repetition of syllables (e.g. *choochoo, night-night*). This possibility is implied by a number of related empirical findings indicating that infants' perceptual and memory constraints are attuned to the processing of repeated elements in strings (Endress, Nespor & Mehler, 2009; Gervain & Werker, 2008). For example, newborns exhibit stronger neural responses to strings of syllables with immediate repetition (e.g. *mubaba*) than those without immediate repetition (e.g. *mubamu, mubage*) (Gervain, Berent & Werker, 2012; Gervain, Macagno, Cogoi, Peña & Mehler, 2008). Six-and-half-month-olds are better at discriminating syllables (e.g. [ba] vs. [du]) when they occur as part of a trisyllabic sequence with two identical syllables (e.g. [**ba**koko] vs. [**du**koko]) rather than two non-identical syllables (e.g. [**ba**tiko] vs. [**du**tiko]), indicating that repeated syllables provide a better context for perceptual discrimination (Goodsitt, Morse, Ver Hoeve & Cowan, 1984). As early as nine months, children are also adept at generalizing patterns between strings that involve repetition of syllables, for example, equating *wo fe fe* with *la di di* as an ABB pattern (Gómez & Gerken, 1999; Marcus, Vijayan, Rao & Vishton, 1999). The advantage of syllable repetition also extends to word learning: eighteen-month-olds learn reduplicated words (e.g. *neenee, foofoo*) more readily than non-reduplicated words (e.g. *neefoo, foonee*) as labels of novel objects (Ota & Skarabela, 2016). As can be seen from the details of the studies mentioned above, however, none of these

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

findings conclusively demonstrate that a sequence consisting of repeated syllables is more likely to be segmented as a word by an infant learner.

It thus remains to be seen whether reduplication facilitates early word segmentation. To address this question, the current study compared infants' segmentation of reduplicated and non-reduplicated word forms from running speech, using a variant of the sequential preferential looking procedure with familiarization and testing (e.g. Ference & Curtin, 2013; Maye, Werker & Gerken, 2002; Thiessen & Erickson, 2013). During familiarization, nine-month-old infants were exposed to two passages, each containing one novel consonant–vowel–consonant–vowel (CVCV) word. One passage contained a novel reduplicated CVCV word (e.g. /nini/) and the other, a novel non-reduplicated CVCV word (e.g. /bole/). During the test phase, we measured the infants' central fixation times for the familiarized words played in isolation. Two previously unheard novel words (one reduplicated, one non-reduplicated) were also added as control items to assess the effects of familiarization. If infants tend to segment reduplicated words over non-reduplicated words, we expect the fixation time difference between the familiarized and non-familiarized items to be larger for the reduplicated words than the non-reduplicated words.

METHOD

Participants

Twenty-four nine-month-olds (13 females, 11 males) were tested. The infants had an average age of 0;8·28 (range: 0;8·12 to 0;9·12). All infants were growing up in the UK, in either a monolingual or predominantly English-speaking family. They were all full-term births and had no reported hearing problems. Eight additional infants were tested but not included due to fussiness/crying (5) and equipment failure (3).

Materials

The critical materials for the experiment consisted of twelve novel words, all with a CVCV structure in English (see Table 1). All words had initial primary stress. In half of them (hereafter 'reduplicated words'), the first and second syllables had an identical segmental composition and in the other half (hereafter 'non-reduplicated words'), the two syllables shared no segments. These words were created by combining three pairs of syllables: *nee* /ni/ and *foo* /fu/; *bo* /bo/ and *lay* /le/; and yah /ja/ and *daw* /dɔ/. This resulted in three sets of reduplicated and non-reduplicated words that had comparable syllable compositions (e.g. *neenee* and *foofoo* on the one hand, versus *neefoo* and *foonee* on the other). To compare other phonological characteristics of the reduplicated versus non-reduplicated words, their phonotactic and neighborhood properties were calculated using a corpus of

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

TABLE 1. *List of experimental words*

Set	Reduplicated	Non-reduplicated
A	*neenee* /nini/	*neefoo* /nifu/
	foofoo /fufu/	*foonee* /funi/
B	*bobo* /bobo/	*bolay* /bole/
	laylay /lele/	*laybo* /lebo/
C	*yahyah* /jaja/	*yahdaw* /jadɔ/
	dawdaw /dɔdɔ/	*dawyah* /dɔja/

NOTE: See the 'Procedure' section for the stimulus assignment method in relation to the experimental word sets.

TABLE 2. *Means (and standard deviations) of the phonotactic and neighborhood properties of the experimental words (all reduplicated versus non-reduplicated comparisons: [$t(10) < 1$])*

	Reduplicated	Non-reduplicated
Positional segment probability	·039 (·015)	·039 (·019)
Biphone probability	·0018 (·0018)	·0022 (·0017)
Number of neighbors	1·00 (2·00)	0·67 (1·21)

English words typically known to children (Storkel & Hoover, 2010). The results, given in Table 2, show that the reduplicated and non-reduplicated words were similar in terms of positional segment probability (the average frequency of each segment in its within-word position), biphone probability (the average frequency of each pair of consecutive segments in the word), and number of neighbors (the number of words in the corpus that differ from the target word by a single sound substitution, deletion, or addition).

Two sets of auditory stimuli were recorded using these experimental words. In the familiarization stimuli, the reduplicated and non-reduplicated words were embedded in the four six-sentence passages used in Jusczyk and Aslin (1995). The real words used in the original study, such as *cup* and *dog*, were replaced by one of the novel words. For example, when the novel reduplicated word *neenee* was embedded in the 'cup' passage from the Jusczyk and Aslin study, the resulting passage was: "The neenee was bright and shiny. A clown drank from the red neenee. The other one picked up the big neenee. [...]". The combination of experimental words and passages was rotated across participants. A complete list of the four carrier passages is given in the 'Appendix'. In the test stimuli, each of the experimental words was read in isolation for fifteen repetitions with varying intonation patterns.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

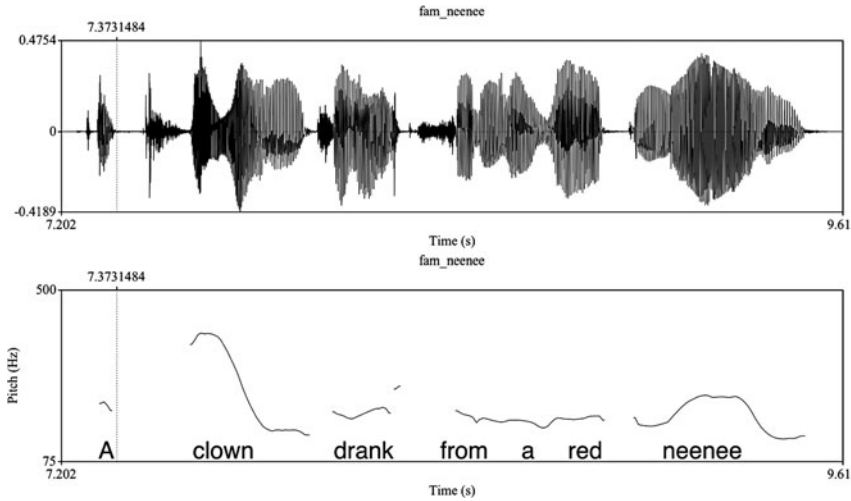


Fig. 1. Waveform and pitch contour of a sample familiarization sentence ("A clown drank from a red neenee").

The familiarization and test stimuli were recorded by a female speaker who was blind to the research question of the study. The reader was encouraged to read the stimuli in a lively voice as if she were talking to an infant. The recordings were made in a soundproof studio, using a Shure SM7 cardioid pattern microphone and a sampling rate of 48 kHz. The familiarization passages were 26·23 s long on average (ranging from 24·84 s to 28·25 s). The mean durations of the passages containing reduplicated versus non-reduplicated words were 26·13 s and 26·33 s, respectively [$t(10) = 0{\cdot}25$, $p = {\cdot}81$]. Figure 1 presents the waveform and pitch contour of a typical sentence in the familiarization passage. All test stimuli were 27 s long, and each began with a 1-s period of silence after which the target test word was repeated fifteen times. In both the familiarization and test stimuli, the reduplicated words and non-reduplicated words were similar in duration, minimum fundamental frequency (F0), maximum F0, and intensity (see Table 3). None of the comparisons between reduplicated versus non-reduplicated words was statistically significant (all ts < 1, n.s.).

Procedure

The experiment was carried out in a dimly-lit sound-attenuated room, equipped with a 47-inch flat-screen TV set to present the visual and audio stimuli, and a remote-controlled digital video camera installed underneath the monitor to record the eye-gaze of the infant. Stimulus presentation

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

TABLE 3. *Means (and standard deviations) of the phonetic properties of the novel word tokens*

	Familiarization		Test	
	Reduplicated	Non-reduplicated	Reduplicated	Non-reduplicated
Duration (ms)	471 (86)	474 (80)	586 (89)	576 (71)
Min Fo (Hz)	157 (32)	156 (29)	131 (8)	131 (8)
Max Fo (Hz)	249 (53)	241 (39)	309 (78)	309 (67)
Intensity (dB)	77·9 (2·2)	77·4 (2·3)	76·8 (1·8)	76·8 (1·7)

was controlled by the Habit 2 program (Oakes, Sperka & Cantrell, 2015). During the experiment, the infant sat on the lap of the caregiver, who was seated approximately 1·5 m away from the display monitor. The caregiver listened to masking music played over headphones and was instructed not to speak to the infant or point at the monitor during the experiment.

The experiment consisted of six familiarization trials and 12 test trials, presented in that order. Trials were interspersed with an attention-getting video sequence that showed colorful moving bubbles with a soundtrack of children's laughter. Each trial was initiated by an experimenter in an adjacent control room, when the experimenter confirmed through a computer monitor that the infant had visually fixated to the attention-getting sequence.

During a familiarization trial, the monitor showed a static image of a white-and-red checkerboard, while one of the familiarization passages was played to completion regardless of the infant's response. Infants were exposed to two passages: one with a reduplicated word and another with a non-reduplicated word. For each infant, the two words were chosen from different sets of the experimental word list (see Table 1) so that there was no segmental overlap between them. For example, an infant who was assigned *neenee* (a Set A word) as a reduplicated word heard a non-reduplicated word from Set B (*bolay* or *laybo*) or Set C (*dawyah* or *yahdaw*). Assignment was counterbalanced across infants. The trials were blocked in groups of two so that each word type occurred once, in random order, in a given block. There were three blocks, resulting in each of the two passages being played three times during the familiarization phase.

During a test trial, the monitor presented a video in which a green circle was shown against a grey background. The circle increased and decreased in size at a 1·28 s cycle. At the same time, one of the auditory test stimuli was played. Trials were infant-controlled. Each trial lasted the full duration of the 27-s stimulus unless the infant looked away for more than two consecutive seconds. The infant's gaze duration and the trial duration were recorded through a key press of the experimenter, who observed the

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

infant on a monitor outside the test room and could not hear the stimulus that was played. To check the reliability of the online measurements, the test trials of three randomly chosen participants (36 trials in total) were coded offline and the duration of central fixation was measured for each trial. The online and offline measurements were highly correlated ($r(34) = \cdot 994$, $p < \cdot 001$). It should be noted, however, that this correlation partly reflects the fact that trial length was infant-controlled and determined by the experimenter's online coding in response to the infant's eye-gaze.

At test, four words were presented to the infant. Two of those were the same reduplicated and non-reduplicated words the infants were exposed to during familiarization. The other two were words the infants had not heard during familiarization, one of which was a reduplicated word and the other one a non-reduplicated word. To avoid any segmental overlap between the four test words, the non-familiar reduplicated word was selected from the same set as the familiar reduplicated word, and the non-familiar non-reduplicated word was selected from a set from which neither one of the familiar words was used. For instance, an infant who was exposed during the familiarization phase to a reduplicated word from Set A (e.g. *neenee*) and a non-reduplicated word from Set B (e.g. *bolay*), received non-familiar test words that included the other reduplicated word from Set A (e.g. *foofoo*) and a non-familiar non-reduplicated word from Set C (e.g. *yahdaw*). As in the familiarization trials, word selection was counterbalanced across infants.

The test trials were blocked in groups of four so that each word type occurred once in a given block. There were three blocks, resulting in each word type being played three times during the test phase. The presentation order within the block was pseudo-randomized to ensure that, across participants, each condition (e.g. familiar non-reduplicated) occurred equally frequently in the first trial, second trial, etc.

RESULTS

Familiarization phase

Mean looking times during the familiarization phase were calculated for each infant. The overall mean looking time was 14·98 s. Across participants, there was no significant difference in the looking times between reduplicated words ($M = 14 \cdot 3$ s, $SD = 4 \cdot 8$) and non-reduplicated words ($M = 15 \cdot 7$ s, $SD = 4 \cdot 6$) [$t(22) = 0 \cdot 75$, $p = \cdot 463$].

Test phase

If infants are more prone to segment reduplicated words than non-reduplicated words in running speech, we should find evidence that, after exposure, they remember familiarized reduplicated words better than

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

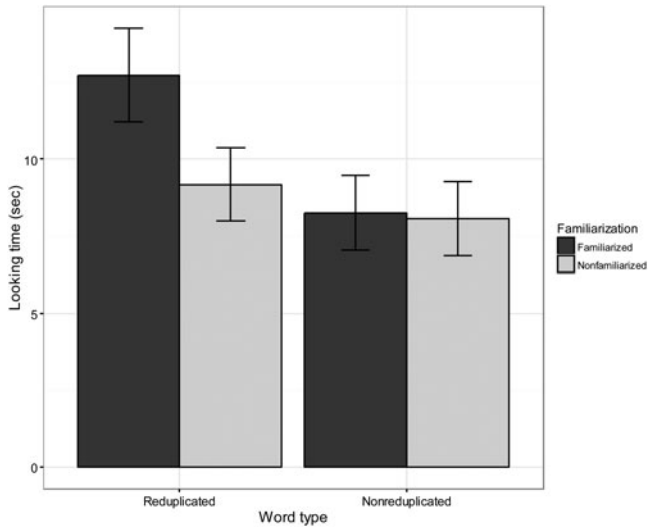


Fig. 2. Mean looking times by word type (reduplicated vs. non-reduplicated) and familiarization (familiarized vs. non-familiarized). Error bars indicate standard errors of mean.

familiarized non-reduplicated words. In the current experiment, this translates to the prediction that infants' performance during the test phase should show larger differentiation in looking times between familiarized and non-familiarized words for the reduplicated words than the non-reduplicated words.

To test this prediction, we measured infants' looking times in the test phase for reduplicated versus non-reduplicated words with or without familiarization. The results are summarized in Figure 2. A 2-way repeated-measures ANOVA with word type (reduplicated vs. non-reduplicated) and familiarization (familiarized vs. non-familiarized) revealed a significant main effect of word type [$F(1,23) = 15{\cdot}56$, $p < {\cdot}001$, $\eta^2 = 0{\cdot}12$], indicating that looking times were generally longer for reduplicated words ($M = 10{\cdot}95$ s, $SD = 4{\cdot}38$) than non-reduplicated words ($M = 8{\cdot}16$ s, $SD = 3{\cdot}66$). The effect of familiarization was also significant [$F(1,23) = 5{\cdot}29$, $p = {\cdot}031$, $\eta^2 = 0{\cdot}06$] due to the overall longer looking times for familiarized words ($M = 10{\cdot}49$ s, $SD = 4{\cdot}84$) than non-familiarized words ($M = 8{\cdot}62$ s, SD 3·36). Crucially, there was also a significant interaction between word type and familiarization [$F(1,23) = 8{\cdot}44$, $p = {\cdot}004$, $\eta^2 = 0{\cdot}05$]. As predicted, the effect of familiarization was larger for reduplicated words (Familiarized: $M = 12{\cdot}73$ s, $SD = 7{\cdot}46$ vs. Non-familiarized: $M = 9{\cdot}17$, $SD = 5{\cdot}79$) than for non-reduplicated words (Familiarized: $M = 8{\cdot}25$ s, $SD = 5{\cdot}94$ vs. Non-familiarized: $M = 8{\cdot}06$, $SD = 5{\cdot}93$).

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

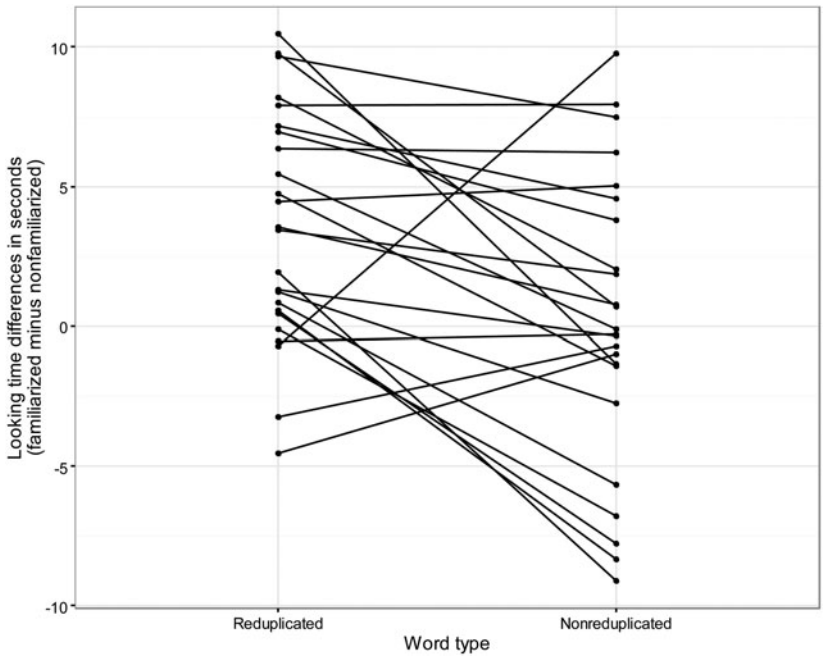


Fig. 3. Effects of familiarization on reduplicated and non-reduplicated words in individual infants. The measure on the *y*-axis is the difference in looking time (s) between the familiarized and non-familiarized condition. Each dot/line represents an infant.

To further examine this interaction, we computed difference scores for each infant by subtracting the mean non-familiarized looking time from the mean familiarized looking time for reduplicated and non-reduplicated words. The results are illustrated in Figure 3. As this figure shows, familiarization had a larger effect on reduplicated than non-reduplicated words in the majority of infants (18 out of the 24). A planned comparison between the difference scores to zero revealed a significant familiarization effect for the reduplicated words [$t(23) = 4{\cdot}16$, $p < {\cdot}001$], but not for the non-reduplicated words [$t(23) < 1$]. The results indicate that the infants remembered the familiarized words only when they were reduplicated.

DISCUSSION

This study demonstrates that reduplication facilitates infants' segmentation of words in continuous speech. The nine-month-olds tested in this study were able to segment novel CVCV words featuring reduplication of syllables (e.g. *neenee*), but not CVCV words without reduplication (e.g. *bolay*). The effect of reduplication revealed in this experiment is striking in

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

light of the recent report that British infants aged eight or nine months, unlike American infants of the same age, do not succeed in word segmentation tasks (Floccia *et al*., 2016). In all the thirteen studies included in Floccia *et al*.'s meta-analysis, however, the target words used in the familiarization passages were monosyllabic or disyllabic real words with no reduplication (e.g. *cup, dog, carriage, kingdom, hamlet, temple*). Consistent with these studies, the British nine-month-olds in the current study failed to learn non-reduplicated words such as *bolay* and *foonee*. Their success in learning the reduplicated novel words highlights the effect of reduplication on their word segmentation.

These findings have important implications for the problem of early word segmentation and further lexical development. Some of the words that infants acquire earliest, such as *papa, daddy, mama*, or *mommy* (Tincoff & Jusczyk, 1999), contain phonological repetition. The current study suggests that the acquisition of such words may be facilitated by a processing bias that preferentially segments strings with repeated elements. These words, in turn, can serve as anchor points for further lexical learning as infants are able to exploit already known words to segment speech input (e.g. Bortfeld *et al*., 2005; Sandoval & Gómez, 2016). In this way, lexical input with phonological repetition offers a bootstrapping mechanism for early word learning.

The results of this study are in line with earlier work showing that repetition of identical units generally holds a special status in infants' processing of linguistic and non-linguistic input (e.g. Endress *et al*., 2009). Crucially, we have presented new evidence that this repetition advantage applies to word segmentation, not only to generalizations of grammatical patterns involving string-internal repetition, which have already been demonstrated (Gerken, Dawson, Chatila & Tenenbaum, 2015; Marcus, Fernandes & Johnson, 2007; Marcus *et al*., 1999; Saffran, Pollack, Seibel & Shkolnik, 2007; see also Gómez & Gerken, 1999). Such effects of phonological repetition on segmentation may explain, at least in part, why infants perform better in a word–object association task when the labels for novel objects are reduplicated, rather than non-reduplicated, words (see Ota & Skarabela, 2016). Our results also shed new light on the findings that infants' perceptual discrimination is better against a phonological context with identical syllables (e.g. [X di di]) than a context with non-identical syllables (e.g. [X di ba]), but only when the stimulus presentation suggests that the syllable strings potentially form words (e.g. with short inter-stimulus intervals) as opposed to sequences of monosyllabic words (Goodsitt, Morgan & Kuhl, 1993; Goodsitt *et al*., 1984). Taken together, these findings indicate that infants' propensity to group and remember the exact tokens of adjacent identical syllables is invoked specifically in the context of learning invariable units, such as words, as opposed to generalized patterns in grammatical structures.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

Infants' preference to segment reduplicated syllables reveals an interesting contrast in the role of repetition in language processing and learning between infants and adults. On the one hand, repetition seems to have similar effects on both infants and adults in the context of pattern generalization. As mentioned earlier, infants as young as nine months are able to generalize structural patterns that involve repetition across different strings (Marcus *et al*., 1999). Adults also benefit from repetition of elements in generalization of structural patterns in artificial grammar learning (Endress, Dahaene-Lambertz & Mehler, 2007; Gómez, Gerken & Schvaneveldt, 2000; Tunney & Altmann, 2001). On the other hand, infants and adults exhibit opposite preferences for repetition in the context of word segmentation. While adults performing online segmentation tasks are sensitive to phonological similarities between syllables in the stimuli (Onnis, Monaghan, Richmond & Chater, 2005), they in fact disprefer strings with repetition of consonants in adjacent syllables (e.g. /popati/) compared to strings without adjacent repetition (e.g. /potipa/) (Boll-Avetisyan & Kager, 2014). The adult preference is consistent with the typological observation that close repetition of identical or similar sounds is avoided in the lexicons of many languages (Monaghan & Zuidema, 2015; Pozdniakov & Segerer, 2007). Thus, while there appears to be an early learning bias that gives rise to a repetition advantage in infants' word segmentation and word learning, mature lexicons are typically not characterized by a tendency for word-internal reduplication.

The contrast between learning biases in infants and the general characteristics of adult language may, however, explain a common feature of infant-directed vocabulary. Although lexical items with reduplication (e.g. *cocoa*) are not common in adult lexicons, they are ubiquitous in the small but unique set of words reserved for the register of infant-directed speech (e.g. *choochoo* and *night-night*; Ferguson, 1964, 1978; see also Endress *et al*., 2007; Gervain & Werker, 2008). Such words may have emerged and been maintained across generations in response to the learning bias demonstrated here. This observation is compatible with theories of language evolution that view language systems as a product of subtle induction biases accumulating over generations of learners via cultural transmission (Christiansen & Ellefson, 2002; Kirby, 2001). Reduplicated words in infant-directed speech may have thus evolved to assist the learner in their discovery of words in continuous speech. This does not of course imply that infants are not able to segment words without reduplication in the input. What the current study adds to the existing research on early word learning is that the range of mechanisms that facilitate word segmentation includes not only learning capacities in infants, but potentially also culturally transmitted properties of the learner's linguistic environment that have been over time shaped by infants' induction biases.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

Our study raises several empirical questions that require further investigations. The familiarization stimuli used in the current study were created with prosodic cues in the experimental words that are strongly in favor of the interpretation that the two syllables are part of the same word rather than repetition of the same word (see Figure 1). If such cues for wordhood are absent, however, it is not clear whether infants will process consecutive identical syllables as one lexical unit with reduplication (e.g. *neenee*) or two repetitions of a monosyllabic lexical unit (e.g. *nee, nee*). To probe this question, it will be necessary to familiarize infants with stimuli devoid of prosodic cues and test whether they segment immediately repeated syllables as a reduplicated disyllabic form or simply as a monosyllabic form. The reduplication effect examined in the current study was also limited to full reduplication, where the entire syllable is repeated. It remains to be seen whether the facilitation effect extends to partial reduplication (e.g. CVCV words with repeated consonants but varied vowels).

In conclusion, the current study demonstrates that reduplication facilitates infants' segmentation of words in continuous speech. Words with repeated syllables may therefore provide a useful entry point for lexical learning. We suggest that this may be one of the functional motivations behind the evolution and maintenance of reduplicated words in the register-unique vocabulary of infant-directed speech across a variety of languages and cultures. The problem of early word segmentation can be tackled not only by general learning strategies, such as the tracking of transitional probabilities, but also by infants' predisposition to detect specific phonological configurations in the input.

REFERENCES

Aslin, R. N., Woodward, J. Z., LaMendola, N. & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (eds), *Signal to syntax: bootstrapping from speech to grammar in early acquisition*, 117–34. Mahwah, NJ: Lawrence Erlbaum.

Boll-Avetisyan, N. & Kager, R. (2014). OCP-PLACE in speech segmentation. *Language and Speech* **57**, 394–421.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M. & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science* **16**, 298–304.

Christiansen, M. H. & Ellefson, M. R. (2002). Linguistic adaptation without linguistic constraints: the role of sequential learning in language evolution. In A. Wray (ed.), *Transitions to language*, 335–58. Oxford: Oxford University Press.

Endress, A., Dehaene-Lambertz, G. & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition* **105**, 577–614.

Endress, A. D., Nespor, M. & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences* **13**, 348–53.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

Ference, J. & Curtin, S. (2013). Attention to lexical stress and early vocabulary growth in 5-month-olds at risk for autism spectrum disorder. *Journal of Experimental Child Psychology* **116**, 891–903.
Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist* **66**, 103–14.
Ferguson, C. A. (1978). Talking to children: a search for universals. In J. Greenberg, C. A. Ferguson & E. A. Moravcsik (eds), *Universals of human language. Volume 1: method and theory*, 203–24. Palo Alto, CA: Stanford University Press.
Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J. & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition* **148**, 1–9.
Gerken, L., Dawson, C., Chatila, R. & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science*, **18**, 80–89.
Gervain, J., Berent, I. & Werker, J. F. (2012). Binding at birth: the newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience* **24**, 564–74.
Gervain, J., Macagno, F., Cogoi, S., Peña, M. & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences* **105**, 14222–27.
Gervain, J. & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass* **2**, 1149–70.
Gómez, R. L. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**, 109–35.
Gómez, R. L., Gerken, L. & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition* **28**, 253–63.
Goodsitt, J. V., Morgan, J. L. & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language* **20**, 229–52.
Goodsitt, J. V., Morse, P. A., Ver Hoeve, J. N. & Cowan, N. (1984). Infant speech recognition in multisyllabic contexts. *Child Development* **55**, 903–10.
Johnson, E. K. & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language* **44**, 548–67.
Jusczyk, P. W. & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* **29**, 1–23.
Jusczyk, P. W., Hohne, E. A. & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics* **61**, 1465–76.
Jusczyk, P. W., Houston, D. M. & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* **39**, 159–207.
Kirby, S. (2001). Spontaneous evolution of linguistic structure–an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* **5**, 102–10.
Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson, (ed.), *Lexical representation and process*, 169–226. Cambridge, MA: MIT Press.
Marcus, G. F., Fernandes, K. J. & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science* **18**, 387–91.
Marcus, G. F., Vijayan, S., Rao, S. B. & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science* **283**, 77–80.
Mattys, S. L., Jusczyk, P. W., Luce, P. A. & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* **38**, 465–94.
Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, B101–11.
Monaghan, P. & Zuidema, W. (2015). General purpose cognitive processing constraints and phonotactic properties of the vocabulary. Paper presented at the *Workshop on the Evolution and Phonetic Capabilities: Causes, Constraints and Consequences*. Retrieved from <https://ai.vub.ac.be/sites/default/files/proceedingsfinal.pdf#page=24>.
Oakes, L. M., Sperka, D. J. & Cantrell, L. (2015). *Habit 2*. Unpublished software. Center for Mind and Brain, University of California, Davis. Online: <habit.ucdavis.edu>.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

Onnis, L., Monaghan, P., Richmond, K. & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language* **53**, 225–37.
Ota, M. & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development* **12**, 380–97.
Pozdniakov, K. & Segerer, G. (2007). Similar place avoidance: a statistical universal. *Linguistic Typology* **11**, 307–48.
Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* **274**, 1926–8.
Saffran, J. R., Pollak, S. D., Seibel, R. L. & Shkolnik, A. (2007). Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition* **105**, 669–80.
Saffran, J. R. & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology* **39**, 484–94.
Sandoval, M. & Gómez, R. L. (2016). Overriding the metrical bias with lexical information: English-learning 7·5-month-olds use *mommy* to segment iambic words. *Language Learning and Development* **12**, 398–412.
Storkel, H. L. & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods* **42**, 497–506.
Thiessen, E. D. & Erickson, L. C. (2013). Discovering words in fluent speech: the contribution of two kinds of statistical information. *Frontiers in Psychology* **3**.
Tincoff, R. & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science* **10**, 172–5.
Tunney, R. J. & Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* **27**, 614–39.

Appendix

Carrier passages used in the familiarization phase (adapted from Jusczyk & Aslin, 1995)

1. The _____ was bright and shiny.
   A clown drank from the red _____.
   The other one picked up the big _____.
   His _____ was filled with milk.
   Meg put her _____ back on the table.
   Some milk from your _____ spilled on the rug.
2. The _____ ran around the yard.
   The postman called to the big _____.
   He patted his _____ on the head.
   The happy red _____ was very friendly.
   Her _____ barked only at squirrels.
   The neighborhood kids played with your _____.
3. The _____ were all different sizes.
   This girl has very big _____.
   Even the toes on her _____ are large.
   The shoes gave the man red _____.
   His _____ get sore from standing all day.
   The doctor wants your _____ to be clean.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press

4. His _____ had big black wheels.
   The girl rode her big _____.
   Her _____ could go very fast.
   The bell on the _____ was really loud.
   The boy had a new red _____.
   Your _____ always stays in the garage.

https://doi.org/10.1017/S0305000916000660 Published online by Cambridge University Press