J. Child Lang. **44** (2017), 1516–1538.
doi:10.1017/S0305000916000623

BRIEF RESEARCH REPORT

Minding the gaps: literacy enhances lexical segmentation in children learning to read*

NAOMI HAVRON AND INBAL ARNON
Hebrew University of Jerusalem, Israel

(*Received* 24 *March* 2016 – *Revised* 28 *August* 2016 – *Accepted* 4 *November* 2016 – *First published online 9 January 2017*)

ABSTRACT

Can emergent literacy impact the size of the linguistic units children attend to? We examined children's ability to segment multiword sequences before and after they learned to read, in order to disentangle the effect of literacy and age on segmentation. We found that early readers were better at segmenting multiword units (after controlling for age, cognitive, and linguistic variables), and that improvement in literacy skills between the two sessions predicted improvement in segmentation abilities. Together, these findings suggest that literacy acquisition, rather than age, enhanced segmentation. We discuss implications for models of language learning.

INTRODUCTION

Two of the major questions in usage-based accounts of language acquisition are (a) How do children extract information from the input they receive? and (b) Why does children's early language differ from that of adults? (Lieven, 2014). The current study touches upon both issues by examining how the experience that older learners have with language, and their knowledge about its structure, impact the units they attend to. We examine how existing knowledge impacts the information children extract from their input by looking at the effect of emergent literacy skills on children's sensitivity to word units in speech. Specifically, we ask how emergent literacy impacts segmentation of multiword units, an investigation that

[*] This work was supported by ISF Grant 52712 (to IA). The authors thank the schools, teachers, parents, and children for their cooperation. We also thank our two anonymous reviewers for their helpful comments. Finally, we thank the research assistants who administered the tasks: Tamar Johnson, Ruth Goldberg, and Yaron Shapira. Address for correspondence: Naomi Havron, Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. e-mail: naomi.havron@mail.huji.ac.il

also aims to add to the relatively scarce literature on the effect of multiword frequency on children (Arnon & Clark, 2011; Bannard & Matthews, 2008; Matthews & Bannard, 2010).

There is recent theoretical interest and empirical support for the idea that MULTIWORD UNITS play an important role in language (Abbot-Smoth & Tomasello, 2006; Arnon, 2010, 2015; Bannard & Lieven, 2009; Biber, 2009; Meunier, 2012; Peters, 1983; Siegelman & Arnon, 2015). We use the term as used in the psycholinguistic literature to refer to multiword sequences produced by speakers (unlike the way the term is used in the formulaic language literature, we do not propose they are stored holistically; Arnon & Snider, 2010). Psycholinguistic findings show that children and adults are sensitive to multiword information in processing (Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Reali & Christiansen, 2007; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010). Data from language acquisition further suggests that children use multiword units in early production and draw on them during learning (Arnon & Clark, 2011; Bannard, Lieven, & Tomasello, 2009; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Salomo, & Tomasello, 2009). However, while both children and adults are sensitive to the properties of multiword units, their role in learning may differ: it has been proposed that children rely more on multiword units during language learning than adult L2 learners, a tendency which impacts learning outcomes (Arnon, 2010; Arnon & Christiansen, 2014; Arnon & Havron, unpublished observations; Arnon & Ramscar, 2012; Siegelman & Arnon, 2015). In this respect, adults' lesser reliance on multiword units may be partially caused by their familiarity with written language where word boundaries are visually salient.

In this paper, we ask if, and how, emergent literacy skills impact the units used during language processing. In particular, we examine if literacy acquisition increases the prominence of single words compared to multiword units. Unlike in spoken language, written language (in many scripts) provides clear information about word boundaries by marking them with spaces. While preliterate children are clearly able to segment utterances into words (Aslin, Saffran, & Newport, 1998; Christophe, Dupoux, Bertonici, & Mehler, 1994), learning to read may nevertheless lead language users to rely more on word units compared to preliterate language users. Such a pattern is consistent with models of language that consider both words and larger units to be building blocks for language (e.g., Bod, 2006; Elman, 2009; McClelland, 2010).

Indeed, studies have shown that literacy increases LEXICAL SEGMENTATION: the ability to segment utterances into words (Gombert, 1994; Havron & Arnon, 2016; Kurvers & Uri, 2006; Olson, 1996; Roberts, 1992; Veldhuis

& Kurvers, 2012). Literate adults and children are more likely to segment sentences into word units compared to illiterate adults or pre-literate children in various tasks, such as counting words in a sentence (Bialystok, 1986), and tapping with chips (Holden & MacGinitie, 1972). However, it is somewhat difficult to separate the effect of age and cognitive and linguistic variables from that of literacy in these studies. Since literate children are typically older than preliterate children, it is possible that the differences in performance are driven by age and the increased linguistic experience that comes with it, rather than by literacy. In line with this interpretation, Karmiloff-Smith and colleagues (Karmiloff-Smith, Grant, Sims, Jones, & Cuckle, 1996) found that five-year-old preliterate children segmented sentences into words more than four-year-old preliterate children, suggesting that age affects segmentation independently of literacy. In contrast, Kurvers, Vallen, and Hout (2006) used a similar task with children in the Netherlands and Norway but did not find an effect of age independent of literacy.

THE CURRENT STUDY

The existing findings leave several open questions. First, the mixed pattern of results on the effect of age highlights the difficulty of separating the effect of age from that of literacy without using longitudinal designs. Second, there is a need to control for other cognitive and linguistic variables which may affect performance on these tasks. Children who can read may differ from their age-matched preliterate peers on additional variables other than literacy: they may have higher cognitive abilities - such as intelligence or short-term memory – and they may have more linguistic knowledge and experience (Evans, Shaw, & Bell, 2000; Gathercole, Tiffany, Briscoe, & Thorn, 2005).

In the current study, we aim to address these questions by examining the effect of emergent literacy on lexical segmentation in preliterate and literate children longitudinally. We will attempt to disentangle the role of literacy and age by using a before-after design, and by measuring and statistically controlling for possible confounding factors: age itself, but also non-verbal intelligence, short-term memory, and vocabulary. We will test Hebrew-speaking first-graders twice: before and after they learn to read. This will allow us to separate the effect of literacy from that of other factors in enhancing lexical segmentation. We test two hypotheses: in the first session, we hypothesize that children's literacy skills will predict segmentation abilities over and above the effects of age and cognitive and linguistic measures. In the second session, we hypothesize that improvement in reading ability will predict improvement in segmentation. If learning to read enhances lexical segmentation, then greater changes in literacy should correspond to larger improvement in segmentation. Importantly, this prediction focuses on changes in literacy between the start and end of the year: we expect the greatest change in segmentation in those children who learned to read during the year (and less so in those who could also read when the year started, as well as those who did not improve and remained poor readers). Note that we do not hypothesize emergent literacy levels at the beginning of the year to predict improvement in segmentation. Such analysis would disregard the fact that children might also improve to a different degree in their literacy skills between the two sessions.

Hebrew is a Semitic language with a rich synthetic morphology. The writing system is an abjad (Daniels & Bright, 1996), i.e., consonants are represented graphically, but vowels are not, for the most part. As such, reading is dependent on knowledge of the morphological elements making up the word – roots, patterns, and the distinction between root consonants and ones that are also used to indicate grammatical morphemes or function affixes (Ravid, 2012). Literacy acquisition in Hebrew and the impact of literacy acquisition on metalinguistic awareness has been studied extensively (see, for example, Bar-On & Ravid, 2011; Berman & Ravid, 2000; Ravid, 2005). Research has focused primarily on the impact of emergent literacy on phonological and morphological awareness (e.g., Levin, Ravid, Rapaport, & Nunes, 1999; Ravid & Malenky, 2001), yet no work to date has explored lexical segmentation in Hebrew. Despite differences between Hebrew and other languages which have been studied in the context of lexical segmentation (Dutch: Kurvers, van Hout, & Vallen, 2009; Onderdelinden, Craats, & Kurvers, 2009; French: Gombert, 1994; Portuguese: Morais, Bertelson, Cary, & Alegrıa, 1986), they share one important feature – they mark word boundaries with spaces. In logographic scripts, such as Chinese Hanzi and Japanese Kanji, there are no conventions for determining word boundaries, and indeed logographic literacy does not seem to aid lexical segmentation (Bassetti, 2007). We thus predict that emergent Hebrew literacy skills will enhance lexical segmentation.

An additional goal of the current study is to add to the sparse literature on multiword processing in children (Arnon & Clark, 2011; Bannard & Matthews, 2008; Matthews & Bannard, 2010). This literature has focused, to date, on the processing of multiword sequences in English. We add to the existing research by asking whether Hebrew-speaking children have a harder time segmenting high-frequency two-word sequences compared to lower-frequency ones. This frequency manipulation will also allow us to validate our task and make sure it is measuring language processing, by replicating frequency effects previously found with adults (e.g., Arnon & Snider, 2010; Bybee, 2002; Havron & Arnon, 2016; Siyanova-Chanturia, Conklin, & van Heuven, 2011). We predict that frequently used two-word sequences will be harder for children to segment, regardless of literacy level.

METHOD

Participants

Thirty-nine children (24 girls and 15 boys) between the ages of five and seven (mean 6;6, range 5;9–7;5) participated in this study. They were recruited from middle- to high-socioeconomic status schools in Israel. Written parental consent was obtained for all children.

MATERIALS

Lexical segmentation tasks

Word order reversal task. In this task, based on a study conducted by Huttenlocher (1964), participants heard two-word sequences and were asked to reverse their order (e.g., hear "little boy", produce "boy little"). The modified task has been used in a study of adult illiterates learning to read in their L2 (Havron & Arnon, 2016). That study showed that task performance was affected by emergent literacy and by the frequency of the two-word sequences, indicating its ability to tap into linguistic processing and not only metalinguistic awareness. All the stimuli were grammatical two-word sequences, used in everyday speech, which differed in the frequency of the two-word sequences but were matched on single-word frequency (e.g., high-frequency *ʔaba ʃeli* 'my dad', low-frequency *tzeva tari* 'fresh paint'). The units included a mixture of more and less fixed expression encompassing a range of grammatical and syntactic relations. The reversed pairs resulted in a mix of grammatical and ungrammatical sequences, as in Huttenlocher's original study (despite the freer word order of Hebrew). We used the MILA Tapuz People Forum Corpus of Israeli online forums (Itai & Wintner, 2008) to calculate the frequencies of the single words and of the two-word sequences. This is a collection of around 1·4 million words taken from Israeli online forums in Hebrew. We used this corpus because (a) there is at the moment no large enough corpus of spoken Hebrew, including child-directed speech (e.g., the Hebrew section of CHILDES contains only about 400,000 tokens) and (b) it is closer to spoken Hebrew that other available written corpora.

We constructed two lists of forty two-word sequences, one for each session. Each list contained an equal number of frequent and infrequent two-word sequences (see 'Appendix 1' for all items). All the two-word sequences were prerecorded by a native speaker of Hebrew unaware of the study goals.[1] The order of presentation was randomized for each participant.

Sentence dictation task. In this task children heard recorded sentences and were asked to say these sentences back to the experimenter "part by part" so that she could write them down. This task is an adaptation of two previously used tasks (Gombert, 1994; Veldhuis & Kurvers, 2012). We controlled for sentence plausibility (as assessed by ratings of native speakers), and the number of words and syllables in the sentences (see 'Appendix 2' for item details). Responses were coded for unit types (word, multiword, and part-word) by two RAs blind to the children's literacy level. Units were defined on the basis of the pauses the child made in her dictation. We define as 'word' a unit that would have been separated by spaces in writing ('teva'). Thus, if the child said: "ima shela [pause] halxa [pause] la@makolet", this is one multiword sequence (ima-shela), and two word units (halxa and la@makolet). We calculated agreement between raters on the number of single words dictated in each sentence, and found reliability to be high (weighted Kappa = $\cdot 89$, $z = 23$, $p < \cdot 001$).

Cognitive and vocabulary measures

Non-verbal intelligence. This was assessed using the child version of the block design subset of the WISC-III (Kaufman, 1994).

Short-term memory. This was evaluated using the digit span task (only forward section; Kaufman, 1994). We used only the digits one to five, to prevent arithmetic ability from affecting performance.

Vocabulary measure. We used the Kavé (Kavé, 2006) Hebrew naming test for productive vocabulary, a validated tool suitable for use with children and adults.

Literacy assessment. This was performed using a rated paragraph reading task. The paragraph was at a level suitable for readers at the end of the first grade. The responses were recorded and later rated by a native speaker who had not met the participants. We used a scale developed for assessing literacy in illiterate populations (Tarone, Biglew, & Hansen, 2007; Havron & Arnon, 2016). The original scale rates both reading and writing, to create a full literacy scale. Here, however, we focus on reading, since it provides children with a visual representation of word boundaries (they are less likely to write full sentences and texts at this stage than read them). Reading was rated for fluency (0–3) and confidence (0–3) by an RA who had not met the children. The scores were aggregated to create a single score (see 'Appendix 3' for paragraphs and rating instructions).

[1] It is possible that having the sequences recorded together as we did (rather than spliced) would lead to more co-articulation cues in the more frequent sequences. However, both types were rated by native speakers as equally natural sounding. Moreover, since factors such as co-articulation also affect real-life language processing, the task may provide a good estimate for online processing.

Twenty percent of children's paragraph readings were coded by a second coder. Inter-rater reliability was high (weighted Kappa = ·9, $z = 2{\cdot}49$, $p = {\cdot}01$).

Current instructions from the ministry of education in Israel determine that children who start first grade should be expected to know the letters and have awareness of the basic speech sounds (phonemic awareness). Children are not expected to be able to read words or sentences. Despite this, many children start school while already being able to read familiar or unfamiliar words, and even short stories. This may be the result of more extensive schooling in a specific kindergarten, or of the children's or their parents' efforts or abilities. For our participants, this resulted in a range of emergent literacy levels, as expanded on in the 'Results' section.

Procedure

Participants were tested twice: once at the start of the school year (October–November), and once after spring break (May), when most children were estimated by the teacher to be able to read aloud a short paragraph without diacritical signs (niqqud). Children were tested in a quiet room using a laptop computer. We used the experimental software E-Prime to run the tasks and record responses. In the first session, participants completed the tasks in the following order: (1) vocabulary, (2) word-order reversal task, (3) block design, (4) sentence dictation task, (5) digit span, and (6) literacy assessment. In the second session the children completed the (1) word-order reversal task, (2) dictation task, and (3) literacy assessment. At the end of each session, the child was given stickers and a certificate of participation.

RESULTS

We excluded participants who completed less than 50% of the trials in the reversal task during the first session, since this was taken to indicate that they did not understand the task. This led to the exclusion of four out of the thirty-nine children who participated in the study.

First session

Literacy, cognitive, and linguistic measures. At the beginning of the first grade, less than half of the children could already read (15 out of 35 children scored 4 and above on the literacy scale). There were no significant differences between the two literacy groups in terms of age, short-term memory, vocabulary, and non-verbal intelligence. As expected, there was a significant difference between the two literacy groups in the reversal task scores, and in the number of word units dictated, with literate children showing more segmentation (see Table 1).

TABLE 1. *Differences between preliterate and literate children at the beginning of the year*

	Preliterate (n = 20)		Literate (n = 15)	
Task	*M*	*SD*	*M*	*SD*
Reversal	81·84%	11·65%	95·64%**	7·16%
Number of words dictated	3·33	1·65	3·96*	1·65
Number of multiword units dictated	1·07	0·79	0·88	0·79
Age	6·55	0·24	6·6	0·4
Digit span	6·19	1·42	5·93	1·71
Vocabulary	32·6	4·19	34·87	4·12
Block design	17·1	9·86	21·8	6·42

NOTES: * $p < ·05$; ** $p < ·01$.

Reversal task. Overall, across both groups, performance ranged between 56·76% and 100%, with a mean of 86·06% ($SD = 11\%$). Cronbach's alpha scores were high (first session, $\alpha = ·87$; second session, $\alpha = ·79$), demonstrating the task's internal reliability. Incorrect responses included repetition of the two-word sequences, content errors (e.g., hearing 'my name', producing *Yoav*), and reversing meaning (e.g., hearing 'little boy', producing 'big girl'). Such errors – which show limited evidence of lexical segmentation – accounted for 73% of errors. Other incorrect responses seem to exhibit some level of segmentation. These included repetition of only one of the words (e.g., hearing 'little boy', producing 'little boy boy'), changing one of the words (e.g., hearing 'little boy', producing 'small boy'), or adding function words (e.g., hearing 'new friend', producing 'friend the new'). Such responses accounted for 27% of the errors.

Hypotheses testing: reversal task. We hypothesized that emergent literacy skills will predict performance after controlling for age, cognitive abilities, and vocabulary. We also hypothesized that more frequent two-word sequences would be harder to segment. To test these predictions, we ran a mixed-effect logistic regression model with trial accuracy as our dependent variable, using the glmer function in R software (Bates, Maechler, Bolker, & Walker, 2015). Since the model predicts the result of each trial we have 1,400 observations in the model from our 35 participants (see Baayen, Davidson, & Bates, 2008; Jaeger, 2008). The dependent variable was accuracy on each trial (as a binary variable). We had logged frequency of two-word sequences and literacy (in *z*-scores) as our variables of interest. We had age (in days), short-term memory, non-verbal intelligence scores, and productive vocabulary scores (all in *z*-scores) as fixed effects. The model included random intercepts for participant and item (the maximal

TABLE 2. *Mixed-effects logistic regression for segmentation task performance: session 1*

	B	SE	Z	p
(Intercept)	4·35	0·46	9·45	<·001
Logged frequency	–0·28	0·12	–2·41	.02
z-score literacy	1·13	0·21	5·31	<·001
z-score age	–0·74	0·25	–2·96	<·001
z-score vocabulary	0·84	0·22	3·80	<·001
z-score non-verbal intelligence	–0·12	0·16	–0·77	0·44
z-score short-term memory	0·78	0·19	4·06	<·001

effect structure that allowed it to converge; Barr, Levy, Scheepers, & Tily, 2013), and had low colinearity (all VIF values < 1·8).

As predicted, the effect of the frequency of two-word sequences on performance was significant: more frequent two-word sequences were harder to reverse (β = –0·28, SE = 0·12, p = ·02, model comparison p = ·02). Importantly, the effect of literacy was also significant: participants who were better readers showed better lexical segmentation (β = 1·13, SE = 0·21, p < ·001, model comparison p < ·001). Surprisingly, age had a negative effect on performance: older children were less likely to produce correct responses after controlling for the other measures (β = 0·74, SE = 0·25, p = ·003, model comparison p = 0·002). Vocabulary size predicted performance, with participants with larger vocabularies performing better (β = 0·84, SE = 0·22, p < ·001, model comparison p < ·001). The effect of short-term memory was also significant: better short-term memory was associated with better performance (β = 0·78, SE = 0·19, p < ·001, model comparison p < ·001). The effect of non-verbal intelligence was not significant (see Table 2).

We suspected that the negative effect of age was driven by the youngest children. For children born between September and December, parents in Israel can choose whether the child will start first grade or remain in kindergarten. It is possible that children whose parents decided to send them to the first grade had characteristics that were correlated both with early literacy and with performance on our task. Put differently, these children may have cognitive or environmental advantages over their older peers. In line with this hypothesis, removing these children (n = 7) eliminated the effect of age (β = –0·41, SE = 0·27, p < ·13) while not impacting the significance of the other factors.

To summarize, as hypothesized, the word order reversal task revealed a significant effect of emergent literacy on lexical segmentation among children of the same age, suggesting that literacy has an independent effect on segmentation abilities. The results also show that children are sensitive

TABLE 3. *Mixed-effects linear model for number of word units in dictation: session 1*

	B	SE	df	t	*p*
(Intercept)	3·60	0·29	11·29	12·51	<·001
z-score literacy	0·31	0·13	28·78	2·48	0·02
z-score age	–0·31	0·13	28·35	–2·36	0·03
z-score vocabulary	0·21	0·13	30·28	1·59	0·12
z-score non-verbal intelligence	–0·01	0·12	29·40	–0·11	0·91
z-score short-term memory	0·23	0·12	28·23	1·88	0·07

to the frequency of two-word sequences and find it harder to segment more frequent two-word sequences.

Dictation task. We looked at the same 35 participants as in the previous analysis. The 330 sentences the participants dictated included 1,191 single words (e.g., *ima* 'mother'), 326 multiword units (e.g., *ima-shela* 'her-mother'), and 37 part words (e.g., *im* 'mothe'). We ran a mixed-effects linear model to predict the number of word units dictated in each sentence (which had a significant and strong negative correlation with the number of multiword units ($r(31) = -0{\cdot}93$, $p < {\cdot}001$). We had literacy as our variable of interest, as well as age (in days), short-term memory, non-verbal intelligence scores, and productive vocabulary scores (all in *z*-scores) as fixed effects. The model included random intercepts for participant and item, and had low colinearity (all VIF values < 1·2).

As predicted, the effect of literacy was significant: participants who were better readers dictated more word units ($\beta = 0{\cdot}31$, SE = 0·13, $p = {\cdot}02$, model comparison $p = {\cdot}01$). Again, age had a significant negative effect on performance: older children were less likely to dictate along word boundaries ($\beta = -0{\cdot}31$, SE = 0·13, $p = {\cdot}03$, model comparison $p = {\cdot}01$). Here again, when we removed children born before December, the effect of age disappeared ($\beta = 0{\cdot}33$, SE = 0·21, $p = {\cdot}13$). Short-term memory had a marginally significant effect ($\beta = 0{\cdot}23$, SE = 0·12, $p = {\cdot}07$, model comparison $p = {\cdot}05$). None of the other factors were significant (see Table 3).

To summarize, as hypothesized, the dictation task also revealed a significant effect of literacy. This task provides convergent evidence to the word order reversal task, showing that emergent literacy, rather than age, is associated with enhanced lexical segmentation.

Second session

Two participants were not tested again at the end of the year; one had left the school and another was unable to complete the word order reversal task for technical reasons. Performance in the word order reversal task ranged

TABLE 4. *Differences in measures between sessions*

Task	Session 1 (n = 33)		Session 2 (n = 33)	
	M	*SD*	*M*	*SD*
Literacy	3	2·38	4·76***	1·57
Reversal task	90·09%	11%	98·08%***	4·17%
Number of words dictated	3·49	1·5	3·81	1·36
Number of multiword dictated	1	0·73	0·8*	0·63

NOTES: * $p < \cdot 05$; *** $p < \cdot 001$.

between 82·5% and 100%, with a mean of 98·08% (SD = 4·17%). There was a significant improvement in reading between the first and second session, as well as in the word order reversal task, and significantly fewer multiword units dictated in the second session (but no significant change in the number of word units; see Table 4).

Hypotheses testing: reversal task. We hypothesized that improvement in literacy would predict improvement in segmentation. To test this prediction, we conducted the following analysis: we first created two improvement scores by regressing first-session literacy scores from second-session literacy scores, and first-session segmentation scores from second-session segmentation scores. We do not use simple improvement scores, but rather, residualized scores, because they are a better estimate for improvement – as they provide scores which are relative to the rest of the group and show whether improvement was above or below expected. Raw improvement scores, in contrast, equate a child who improved from 5 to 6 (as expected) with a child who improved from 0 to 1 (much less than expected). After creating these improvement scores, we conducted a linear regression, predicting segmentation improvement score (in *z*-scores) from literacy improvement scores (in *z*-scores). As expected, we find that literacy acquisition is significantly correlated with segmentation enhancement ($R = \cdot 02$, $t(31) = 3\cdot 15$, $p < \cdot 004$). Improvement in literacy also explained a significant proportion of variance in segmentation improvement scores (corrected $R^2 = \cdot 22$) (see Figure 1). Since participants who were already reading fluently in the first session and had very high segmentation accuracy could not have improved much in either task, their performance may support our hypothesis for psychometric and not substantial reasons (they had little room for improvement on either task). Thus, to show that our effect is not driven solely by these participants, we conducted a second analysis, in which we removed all participants who were at the top quartile of scores on lexical segmentation in the first session, and who could also read at level 6. Even though removing these children left us with a smaller sample (n = 18) the findings remain the

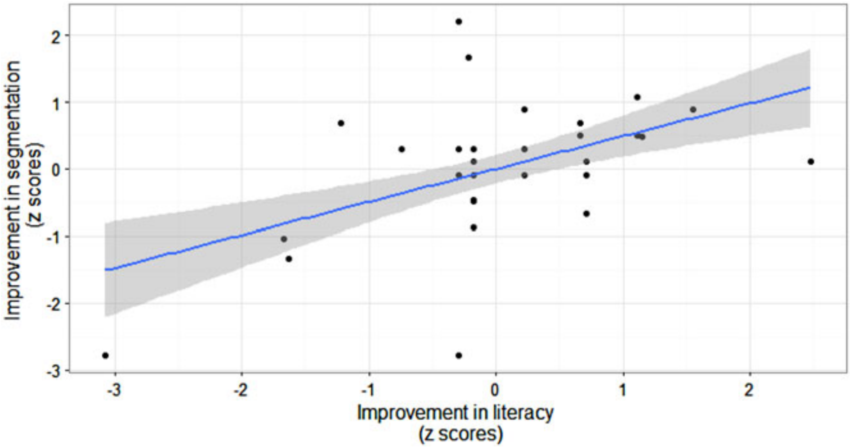

Fig. 1. Improvement in segmentation by improvement in literacy.

same: improvement in literacy is significantly correlated with improvement in segmentation ($R = {\cdot}02$, $t(16) = 2{\cdot}98$, $p < {\cdot}01$). Moreover, improvement in literacy explained a larger proportion of variance in segmentation improvement scores than in the analyses that included all children (corrected $R^2 = {\cdot}32$). To conclude, the more a child improved in her literacy skills, the greater improvement she showed in the reversal task.

Dictation task. We hypothesized that improvement in literacy would predict an increase in the number of word units dictated (which will serve as evidence for segmentation into word units). To test this prediction, we conducted the same analysis as we performed for the reversal task: we first created two improvement scores by regressing first-session literacy scores from second-session literacy scores, and the mean number of words the child dictated per sentence in the first session from the mean number of words the child dictated per sentence in the second session. After creating these improvement scores, we conducted a linear regression, predicting increase in number of words dictated per sentence (in z-scores) from improvement score in literacy (in z-scores). We find that literacy acquisition marginally significantly predicts the increase in the number of word units ($R = {\cdot}34$, $t(31) = 1{\cdot}85$, $p = {\cdot}07$). As with the reversal task, we ran an additional regression after removing participants who were already reading at level 6 at the first session (here there is no potential of a ceiling effect in the dictation task). The analysis now yields a significant effect ($R = {\cdot}35$, $t(23) = 2{\cdot}31$, $p = {\cdot}03$). Improvement in literacy explains a small proportion of variance in segmentation improvement scores (corrected $R^2 = {\cdot}16$).

To summarize, both tasks show evidence that acquisition of literacy is correlated with improvement in segmentation.

DISCUSSION

The present study had two major goals: to examine the effect of emergent literacy skills on lexical segmentation, and to add to the literature on child processing of multiword sequences. Regarding the first goal, we found strong evidence that emergent literacy skills affect lexical segmentation after controlling for cognitive and linguistic abilities. Somewhat surprisingly, we found a negative correlation between age and segmentation after controlling for literacy. We believe this finding resulted from factors confounded with early literacy skills that were not directly tested in the current study. In particular, children who can read before first grade may differ from preliterate children of the same age in many respects. While we controlled for some cognitive and linguistic variables, we did not examine other potentially confounding factors such as possible differences in parental education and socioeconomic status, literacy readiness training in kindergarten, phonological awareness, etc., all of which may impact both emergent literacy skills and language processing (de Jong & van der Leij, 1999). Indeed, removing the children who could have stayed in kindergarten an extra year eliminated the effect. Our control of various factors related to language processing and experience (age, vocabulary, short-term memory, non-verbal intelligence), and the fact that the size of improvement in the segmentation tasks was related to the size of improvement in literacy, are a strong indication that emergent literacy plays a causal role in the development of lexical segmentation. While it is possible that both improvement in segmentation and improvement in literacy are the result of a third, confounding, factor, the combined findings from the two sessions provide evidence that literacy indeed enhances segmentation. This joins the large number of studies which point to the role of literacy in enhancing segmentation (e.g., Gombert, 1994; Kurvers & Uri, 2006; Roberts, 1992; Veldhuis & Kurvers, 2012), while adding the control of age, cognitive abilities, and the use of a more online task. However, one limitation of the study is that, though less metalinguistic than previously used tasks, and although they do replicate findings from the online psycholinguistic literature (multiword frequency effects), our tasks are still quite offline.

Regarding our second goal, we found that frequency had a negative effect on performance in the word order reversal task at the beginning of the year. The more frequent the two-word sequence, the harder it was for children to reverse, mirroring adult studies on multiword frequency effects (e.g., Arnon & Snider, 2010; Bybee, 2002; Havron & Arnon, 2016; Reali & Christiansen, 2007; Tremblay & Baayen, 2010). We suggest that the children found it hard to override frequency-induced word-order preferences for high-frequency two-word sequences (e.g., Siyanova-Chanturia *et al.*, 2011). We do not find it likely that this effect was driven by the fact that children do not know that the two-word sequences consist of two separate words. At least in the overwhelming majority of cases, we believe that children did know that the bigrams consist of two different words, given their age and the nature of the stimuli, which included high-frequency and familiar words (e.g., 'good night', or 'each one'). Our study joins the few studies conducted on children which show that children display the same multiword frequency effects as adults (Arnon & Clark, 2011; Bannard & Matthews, 2008; Matthews & Bannard, 2010). It is also the first study to show such effects in Hebrew-speaking children.

Our findings may have consequence for models of language learning. Multiword units are seen as important building blocks for first language learning in usage-based approaches (e.g., Abbot-Smith & Tomasello, 2006; Arnon & Christiansen, 2014; Bod, 2006). It has also been suggested that some aspects of grammar are better learned from multiword units than from single words (Arnon, 2010; McCauely & Christiansen, 2011, 2014; Wray, 2000). Empirical support for these suggestions comes from artificial language learning experiments, which found that learning from utterances was more beneficial for learning grammatical relations between words, compared to learning from smaller units (Arnon, 2010; Arnon & Ramscar, 2012; Siegelman & Arnon, 2015; Valian & Levit, 1996). If emergent literacy skills also impact the units of learning, and lead to reduced focus on multiword units, then preliterate children may show an advantage over literate children in learning these relations. In fact, we have recently completed an artificial language study with preliterate and literate children, which found evidence that emergent literacy skills also affect language learning (Havron & Arnon, unpublished observations).

To conclude, we showed that emergent literacy skills affect lexical segmentation, even when controlling for age, cognitive and linguistic variability, and using a relatively online task. We also showed an improvement in lexical segmentation after literacy acquisition, with the amount of improvement in literacy predicting enhancement in segmentation abilities, suggesting that it is literacy, rather than age, which leads to enhanced segmentation. We add to the growing literature on the role of multiword units in language processing in children, and expand pervious research to lexical segmentation in Hebrew.

REFERENCES

Abbot-Smith, K. & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review* **23**(3), 275–290.

Arnon, I. (2010). *Starting big: the role of multiword phrases in language learning and use*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
Arnon, I. (2015). What can frequency effects tell us about the building blocks and mechanisms of language learning? *Journal of Child Language* **42**, 274–277.
Arnon, I. & Christiansen, M. H. (2014). Chunk-based language acquisition. In P. J. Brooks & V. Kempe (eds), *Encyclopedia of language development*, 88–90. Thousand Oaks, CA: Sage Publications.
Arnon, I. & Clark, E. V. (2011). Why brush your teeth is better than teeth – children's word production is facilitated in familiar sentence-frames. *Language Learning and Development* **7**, 107–129.
Arnon, I. & Cohen Priva, U. (2013). More than words: the effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* **56**(3), 349–371.
Arnon, I. & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition* **122**(3), 292–305.
Arnon, I. & Snider, N. (2010). More than words: frequency effects for multi-word phrases. *Journal of Memory and Language* **62**(1), 67–82.
Aslin, R. N., Saffran, J. R. & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* **9**(4), 321–324.
Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390–412.
Bannard, C. & Lieven, E. (2009). Repetition and reuse in child language learning. *Formulaic Language* **2**, 299–321.
Bannard, C., Lieven, E. & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences of the United States of America* **106**(41), 17284–17289.
Bannard, C. & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science* **19**, 241–248.
Bar-On, A. & Ravid, D. (2011). Morphological analysis in learning to read pseudowords in Hebrew. *Applied Psycholinguistics* **32**(3), 553–581.
Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**(3), 255–278.
Bassetti, B. (2007). Bilingualism, biliteracy and metalinguistic awareness: word awareness in English and Japanese users of Chinese as a second language. *Birkbeck Studies in Applied Linguistics* **2**, 1–21.
Berman, R. A. & Ravid, D. (2000). Acquisition of Israeli Hebrew and Palestinian Arabic: a review of current research. *Hebrew Studies* **41**, 83–98.
Bialystok, E. (1986). Children's concept of word. *Journal of Psycholinguistic Research* **15**(1), 13–32.
Biber, D. (2009). A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* **14**, 275–311.
Bod, R. (2006). Exemplar-based syntax: how to get productivity from examples. *Linguistic Review* **23**(3), 291–320.
Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. <http://CRAN.R-project.org/package=lme4>.
Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* **24**(2), 215–221.
Christophe, A., Dupoux, E., Bertonici, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America* **95**(3), 1570–1580.
Daniels, P. T. & Bright, W. (eds.) (1996). *The world's writing systems*. New York: Oxford University Press.

de Jong, P. F. & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology* **91**(3), 450–476.
Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science* **33**(4), 547–582.
Evans, M. A. N. N., Shaw, D. & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale* **54**(2), 65–75.
Gathercole, S. E., Tiffany, C., Briscoe, J. & Thorn, A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **46**(6), 598–611.
Gombert, J. E. (1994). How do illiterate adults react to metalinguistic training? *Annals of Dyslexia* **44**, 250–269.
Havron, N. & Arnon, I. (2016). Reading between the words: the effect of literacy on second language lexical segmentation. *Applied Psycholinguistics*. Online: <http://dx.doi.org/10.1017/S0142716416000138>.
Holden, M. H. & MacGinitie, W. H. (1972). Children's conceptions of word boundaries in speech and print. *Journal of Educational Psychology* **63**(6), 551–557.
Huttenlocher, J. (1964). Children's language: word–phrase relationship, **143**(3603), 264–265.
Itai, A. & Wintner, S. (2008). Language resources for Hebrew. *Language Resources and Evaluation* **42**(1), 75–98.
Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* **59**, 434–446.
Karmiloff-Smith, A., Grant, J., Sims, K., Jones, M. C. & Cuckle, P. (1996). Rethinking metalinguistic awareness: representing and accessing knowledge about what counts as a word. *Cognition* **58**(2), 197–219.
Kaufman, A. S. (1994). *Wiley series on personality processes: intelligent testing with the WISC-III*. Oxford: John Wiley.
Kavé, G. (2006). The development of naming and word fluency: evidence from Hebrew-speaking children between ages 8 and 17. *Developmental Neuropsychology* **29**(3), 493–508.
Kurvers, J. & Uri, H. (2006). Metalexical awareness: development, methodology or written language? A cross-linguistic comparison. *Journal of Psycholinguistic Research* **35**(4), 353–367.
Kurvers, J., Vallen, T. & Hout, R. V. (2006). Discovering features of language: metalinguistic awareness of adult illiterates. *LOT Occasional Series* **6**, 69–88.
Kurvers, J., van Hout, R. & Vallen, T. (2009). Print awareness of adult illiterates: a comparison with young pre-readers and low-educated adult readers. *Reading and Writing* **22**, 863–887.
Levin, I., Ravid, D., Rapaport, S. & Nunes, T. (1999). Developing morphological awareness and learning to write: a two-way street. *Neuropsychology and Cognition* **17**, 77–104.
Lieven, E. (2014). First language development: a usage-based perspective on past and current research. *Journal of Child Language* **41**(1), 48–63.
Lieven, E., Behrens, H., Speares, J. & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language* **30**(2), 333–370.
Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: a usage-based analysis. *Cognitive Linguistics* **20**(3), 481–507.
Matthews, D. & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science* **34**(3), 465–488.
McCauley, S. M. & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: the CAPPUCCINO model. In L. Carlson, C. Hölscher & T. Shipley (eds), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1619–1624). Austin, TX: Cognitive Science Society.

McCauley, S. M. & Christiansen, M. H. (2014). Acquiring formulaic language: a computational model. *Mental Lexicon* **9**(3), 419–436.
McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science* **2**(4), 751–770.
Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics* **32**, 111–129.
Morais, J., Bertelson, P., Cary, L. & Alegrıa, J. (1986). Literacy training and speech segmentation. *Cognition* **24**, 45–64.
Olson, D. R. (1996). Towards a psychology of literacy: on the relations between speech and writing. *Cognition* **60**(1), 83–104.
Onderdelinden, L., Craats, I. V. D. & Kurvers, J. (2009). Word concept of illiterates and low-literates: worlds apart? *LOT Occasional Series* **15**, 35–48.
Peters, A. M. (1983). *The units of language acquisition*. New York: Cambridge University Press.
Ravid, D. (2005). Hebrew orthography and literacy. In R. Malatesha Joshi & P. G. Aaron (eds), *Handbook of orthography and literacy*, 339–363. Mahwah, NJ: Lawrence Erlbaum Associates.
Ravid, D. (2012). *Spelling morphology: the psycholinguistics of Hebrew spelling*. New York: Springer.
Ravid, D. & Malenky, A. (2001). Awareness of linear and nonlinear morphology in Hebrew: a developmental study. *First Language* **21**(61), 25–56.
Reali, F. & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* **57**, 1–23.
Roberts, B. (1992). The evolution of the young child's concept of 'word' as a unit of spoken and written language. *Reading Research Quarterly* **27**(2), 125–138.
Siegelman, N. & Arnon, I. (2015). The advantage of starting big: learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language* **85**, 60–75.
Siyanova-Chanturia, A., Conklin, K. & van Heuven, W. J. B. (2011). Seeing a phrase 'time and again' matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition* **37**(3), 776–784.
Tarone, E., Bigelow, M. & Hansen, K. (2007). The impact of alphabetic print literacy level on oral second language acquisition. In N. Faux (ed.), *Low-educated second language and literacy acquisition: research, policy, and practice. Proceedings of the second annual forum of the Literacy Institute*, 99–122. Richmond, VA: Virginia Commonwealth University, Literacy Institute.
Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: a behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (ed.), *Perspectives on formulaic language: acquisition and communication*, 151–173. London: Continuum International Publishing Group.
Valian, V. & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *Journal of Memory and Language* **35**(4), 497–516.
Veldhuis, D. & Kurvers, J. (2012). Offline segmentation and online language processing units: the influence of literacy. *Written Language & Literacy* **15**, 165–184.
Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* **21**, 463–489.

Appendix 1: stimuli for word pair reversal task

Session	Hebrew two-word sequences	Translation	Type	Frequency	Word 1 frequency	Word 2 frequency	Plausibility	# syllables
1	ʔein tsorex	there is no need	freq	53	2331	151	3·08	3
1	ʔanaʃim raʔim	bad people	freq	31	1710	128	2·80	5
1	xag ʃmeax	happy holiday	freq	72	176	176	1·85	4
1	jihije tov	it will be good	freq	344	1275	2855	3·97	4
1	jom huledet	birthday	freq	45	935	88	1·70	4
1	joter miday	too much	freq	75	4846	196	1·60	4
1	Y = jaxol lihijot	could be	freq	252	1413	1803	3·20	5
1	ma ʔosim	what do we do	freq	54	4792	377	3·70	3
1	ma kara	what's up	freq	53	4792	211	2·70	3
1	paʔam riʃona	first time	freq	37	1162	132	1·58	5
1	bli keʃer	regardless	freq	22	700	393	1·60	3
1	im yeʃ	if there is	freq	287	4879	4788	3·10	2
1	ʔal tidagi	don't worry	freq	44	1126	48	1·41	4
1	derex xaim	way of life	freq	162	780	1420	2·61	4
1	kos sukar	a cup of sugar	freq	72	621	383	1·57	3
1	kol ʔexad	each one	freq	424	5730	2198	1·26	3
1	layla tov	good night	freq	32	107	2855	2·20	3
1	mits limon	lemon juice	freq	63	139	138	1·38	3
1	ʔad kama	to what extent	freq	108	2049	1785	2·69	3
1	toda raba	thanks very much	freq	253	1653	348	1·50	4
1	ʔeize jelled	what child	infreq	3	895	250	1·70	4
1	ʔeix tsarix	how one needs	infreq	2	1346	1208	1·90	3
1	panter ʃaxor	black panther	infreq	1	9	264	1·88	4
1	gan gadol	big kindergarten	infreq	1	182	565	1·56	3
1	lo xazarti	I didn't come back	infreq	2	15342	423	2·00	4
1	ʔamur ladaat	should know	infreq	2	229	337	1·70	5
1	xaver ʔexad	one friend	infreq	1	230	2198	2·40	4
1	ʔoto gova	same hight	infreq	2	1664	46	2·20	4
1	basar kafu	frozen meat	infreq	2	178	21	1·20	4

Appendix 1: (*cont.*)

Session	Hebrew two-word sequences	Translation	Type	Frequency	Word 1 frequency	Word 2 frequency	Plausibility	# syllables
1	liʃtof ʔotam	wash them	infreq	1	58	665	2·60	4
1	ʔadam ʃone	different person	infreq	1	802	151	2·00	4
1	ʔorez ragil	regular rice	infreq	1	176	126	1·40	4
1	geʃem kal	light rain	infreq	2	75	443	1·86	3
1	xaxam gadol	so clever	infreq	2	51	565	2·70	4
1	kapit ktana	small teaspoon	infreq	1	379	311	1·41	4
1	ʔuga metuka	sweet cake	infreq	1	102	141	1·91	5
1	ʔets jarok	green tree	infreq	2	154	123	2·37	3
1	ʔaʃan ʃaxor	black smoke	infreq	2	10	264	2·25	4
1	tseva tari	fresh paint	infreq	3	71	147	1·13	4
1	ʃana ʔaxat	one year	infreq	3	965	893	1·53	4
2	ʔeix ʔefʃar	how can one	freq	45	1346	1670	1·85	3
2	ʔein beʔaja	no problem	freq	21	2331	272	2·90	5
2	gvina levana	spread cheese	freq	28	168	54	1·30	5
2	gan xajot	zoo	freq	24	182	198	2·70	3
2	xatsi ʃana	half a year	freq	49	478	965	1·20	4
2	misxak kal	easy game	freq	61	552	443	1·50	3
2	paʔam ʔaxat	one time	freq	50	1162	893	1·60	4
2	gan ʔeden	Heaven	freq	46	182	59	1·18	3
2	ʔofen haxana	way of preparation	freq	53	379	144	1·44	5
2	davar kal	easy thing	freq	43	1001	443	2·00	3
2	haxi tov	best	freq	347	1230	2855	1·89	3
2	xatsi kos	half a cup	freq	75	478	621	1·38	3
2	kol nose	every subject	freq	173	5730	269	2·38	3
2	maim rotxim	hot water	freq	51	457	68	1·93	4
2	mikre ʔavud	lost cause	freq	58	344	65	1·44	4
2	kodem kol	first of all	freq	133	281	5730	1·82	3
2	resek ʔagvanijot	tomato sauce	freq	62	91	199	1·59	5
2	ʃalom rav	[big] hello	freq	42	787	191	2·06	3
2	ʃemen zajit	olive oil	freq	131	523	181	1·93	4

2	ʃana tova	good year	freq	278	965	752	1·44	4
2	ʔaz kama	so how many	infreq	3	2870	1785	1·50	3
2	joter ʔor	more light	infreq	2	4846	232	2·26	3
2	jaxol lexapes	can look for	infreq	2	1413	159	2·34	5
2	lo xadaʃ	not new	infreq	2	15342	423	1·78	3
2	ma niʃʔar	what is left	infreq	3	4792	203	3·20	3
2	ma sheli	what is mine	infreq	2	4792	2686	2·20	3
2	ʔim horim	with parents	infreq	3	6550	349	3·40	3
2	boker xadaʃ	new morning	infreq	1	270	423	1·50	4
2	mamaʃ hamon	really a lot	infreq	2	1481	494	3·30	4
2	meʔanyenʔ oto	interests him	infreq	1	1664	285	3·90	5
2	xibuk gadol	big hug	infreq	2	26	2198	2·35	4
2	ʔover leʔat	passes slowly	infreq	1	104	93	2·68	4
2	ʔerev ʃabat	shabat evening	infreq	3	143	240	2·70	4
2	ʔeser dakot	ten minutes	infreq	2	1101	790	1·61	4
2	ʃavua xadaʃ	new week	infreq	3	331	423	1·21	5
2	ʃulxan ʔavoda	work desk	infreq	1	29	404	1·43	5
2	toxnit xadaʃa	new program	infreq	3	115	229	1·28	5
2	tmuna yafa	pretty picture	infreq	1	242	416	2·22	4
2	tapuʔax ʔadom	red apple	infreq	1	69	181	2·62	5
2	tʃuva nexona	correct response	infreq	3	112	128	1·75	5

Appendix 2: stimuli for dictation task

Session	Sentence in Hebrew	English translation	# words	# syllables	Plausibility
1	hem tamid ʔoxlim gvina levana ba@ʔerev	They always eat cheese at the evening	6	14	1·20
1	ʔetmol raʔinu gan xayot gadol meʔod	Yesterday we saw a really big zoo	6	12	1·87
1	ʔani ʔixalti lo xag sameʔax be@pesax	I wished him a happy holiday in Passover	6	12	2·41
1	ʔein li joter midai ma lidʔog	I don't really have anything to worry about	6	9	1·30
1	hu lo jadʔa ma laʔasot ʔita	He didn't know what to do with her	6	10	1·40
1	hem lo jodʔim ma ʔosim ʔaxsʃav	They don't know what to do next	6	9	1·45
1	hem sixaku ʔod paam ba@misxak ha@ze	They played that game again	6	12	2·23
1	hi halxa la@rofe ʔim ʔima ʃela	She went to the doctor's with her mother	6	14	1·05
1	zot lo paʔam riʃʔona ʃeli kʔan	This is not my first time here	6	10	1·68
1	yaʔel tsrixa lihijot ba@bait ʔaxʃav	Yael need to be home right now	5	11	1·17
2	ʃe@ihiye lexa jom huledet sameax	Have a happy birthday	5	15	1·55
2	dana kol jom shishi maxina ʔugiyot	Dana makes cookies every Friday	6	14	1·71
2	ha@ben ʃeli kaze jeled katan ve@xamud	My son is such a cute and little boy	6	14	1·90
2	ʔani yodeʔa ʔeifo jeʃ kʔan ʃerutim	I know where there is a bathroom here	6	12	1·37
2	le@sara yeʃ kvar ʃloʃa nexadim	Sara already has three grandchildren	5	12	1·15
2	hem lomdim kol jom ba@ʃavua ha@ba	They study every days next week	8	17	1·50
2	ʃavua ʔexad ze loʔ harbe zman	One week is not a long time	6	15	1·65
2	ʔani kvar lo mexapes dira	I am not looking for an apartment anymore	5	15	1·40
2	ʔani lo mevin ma tsarix laʔasot	I don't understand what I have to do	6	16	1·10
2	ʔani roze ladaʔat ma kara ʔetmol	I want to know what happened last week	6	14	1·87

NOTES: @ marks a prefix, which we coded as part of the next word, since in Hebrew there is no space between the prefix and the word.

Appendix 3: reading scale

Session 1:

שַׁבת זה יום של משפחה.

מי בבית בשבת?

אבא, אמא, תמי, יוסי.

סבא וסבתא באים לבקר.

יחד בסלון כלם.

ואני גם.

English translation: Saturday is a day for families. Who is home on Saturday? Dad, Mom, Tami, Yossi. Grandpa, and grandma come to visit. Everyone is in the living room together. So am I.

Session 2:

פעם היה עץ אורן.

העץ רצה חבר.

בא השמש, בא ענן,

באה יונה,

והעץ לא מצא חבר.

בא דני ושתל עץ קטן

על יד האורן.

היה חבר,

לא היה לבד.

English translation: Once there was a pine tree. The tree wanted a friend. The sun came, the cloud came, the dove came, and the tree could not find a friend. Danny came, and planted a small tree next to the pine. The tree had a friend, the tree was not alone.

Coding instructions
Fluency

0	Did not read the paragraph.
1	Read slowly, backtracked, did not understand what he is reading (this is sometimes obvious from how the experimenter is reacting), asked for help from the experimenter. Read words incorrectly and did not try to correct himself.
2	Started slowly but picked up the pace. Still found it hard to decipher word meanings. Sometimes read the same thing twice, faster the second time.
3	Read comfortably, relatively fast, little difficulty in understanding.

Confidence

0	Did not read the paragraph.
1	Said he does not want to or cannot read (but still agreed to try). If a child did not finish reading the paragraph he will also be given 1, even if he did not hesitate to try.
2	Tried, but was uncertain of his abilities. Asked questions throughout the process. Tried and read (not fluently), but showed no confidence.
3	Approached the paragraph reading without hesitation.