

L2 Processing Advantages of Multiword Sequences: Evidence from Eye-Tracking

Elma Kerz

RWTH Aachen University

elma.kerz@

ifaar.rwth-aachen.de

Arndt Heilmann

RWTH Aachen University

arndt.heilmann@

ifaar.rwth-aachen.de

Stella Neumann

RWTH Aachen University

stella.neumann@

ifaar.rwth-aachen.de

Abstract

A substantial body of research has demonstrated that native speakers are sensitive to the frequencies of multiword sequences (MWS). Here, we ask whether and to what extent intermediate-advanced L2 speakers of English can also develop the sensitivity to the statistics of MWS. To this end, we aimed to replicate the MWS frequency effects found for adult native language speakers based on evidence from self-paced reading and sentence recall tasks in an ecologically more valid eye-tracking study. L2 speakers' sensitivity to MWS frequency was evaluated using generalized linear mixed-effects regression with separate models fitted for each of the four dependent measures. Mixed-effects modeling revealed significantly faster processing of sentences containing MWS compared to sentences containing equivalent control items across all eye-tracking measures. Taken together, these findings suggest that, in line with emergentist approaches, MWS are important building blocks of language and that similar mechanisms underlie both native and non-native language processing.

1 Introduction

1.1 Emergentist approaches and statistical learning

A widely held assumption in the language sciences, including psycholinguistics, has long been the 'words and rules' view (Levelt, 1993; Jackendoff and Jackendoff, 2002; Pinker, 1999): In this view, speakers/writers generate sentences by combining words according to the grammatical rules of their language, and listeners/readers comprehend sentences by looking up words in their mental lexicon and combining them using the same rules. This view has been challenged recently by an accumulating body of evidence demonstrating that language users are highly sensitive not only to

the frequencies of individual words but also to the frequencies of word sequences (see, e.g., Christiansen and Arnon, 2017, for a recent overview). This questions the strict compartmentalization between the lexicon as a storage of individual words and a grammar as a set of rules or constrained used to combine them.

Moving away from the traditional 'words and rules' approach, emergentist approaches have put forward alternative theoretical models of language. Following the literature (see, e.g. Arnon and Snider, 2010; Kidd et al., 2017; MacWhinney and O'Grady, 2015; Mitchell et al., 2013), we use the term 'emergentist' as a cover term for a broad class of approaches to language including usage-based (a.k.a. experience-based) models, constraint-based approaches, exemplar-based models and connectionist models (for more general overviews, see, e.g., Beckner et al., 2009; Christiansen and Chater, 2016a,b; Ellis and Larsen-Freeman, 2006; Ellis, 2019; MacWhinney, 2012; McClelland et al., 2010). Distinct from nativist/generative approaches, emergentist approaches share the following two central assumptions: First, emergentist approaches eschew the existence of Universal Grammar and instead emphasize that language is learnable via general cognitive mechanisms. Second, these approaches put the emphasis on usage and/or experience with language and assume a direct and immediate relationship between processing and learning, conceiving of them as inseparable rather than governed by different mechanisms ('two sides of the same coin'). In these approaches, language acquisition is viewed as learning how to process efficiently (see, the 'learning-as-processing' assumption, Chang, Dell, and Bock, 2006; see also 'language acquisition as skill learning' Chater and Christiansen, 2018). One of the major advances in the language sciences across theoretical orienta-

tions has been the recognition that language consists of complex, highly variable patterns occurring in sequence, and as such can be described in terms of statistical or distributional relations among language units (see, e.g., Redington and Chater, 1997). Thus, learning a language heavily involves figuring out the statistics inherent in language input. This is supported by a large body of evidence from the literature on statistical learning. Statistical learning – defined as the mechanism by which language users discover the patterns inherent in the language input based on its distributional properties – has been shown to facilitate the acquisition of various aspects of language knowledge, including phonological learning (e.g., Maye et al., 2008; Thiessen and Saffran, 2003), word segmentation (e.g., Onnis et al., 2008; Saffran et al., 1996), learning the graphotactic and morphological regularities of written words (e.g., Pacton et al., 2005), learning to form syntactic and semantic categories and structures (e.g., Lany and Saffran, 2010; Saffran and Wilson, 2003; Thompson and Newport, 2007). Furthermore, an impressive body of evidence has been accumulating over the last years indicating a close relationship between individual differences in statistical learning ability and variation in native language learning in both child and adult L1 populations (e.g., Conway et al., 2010; Kidd and Arciuli, 2016; Misyak and Christiansen, 2012; Siegelman and Frost, 2015), and in adult L2 populations (e.g., Ettliger et al., 2016; Frost et al., 2013; Onnis et al., 2016). Thus, from an emergentist perspective, language acquisition is essentially an ‘intuitive statistical learning problem’ (Ellis, 2008, p. 376).

Emergentist approaches have developed a growing interest in the role of multiword sequences (henceforth MWS), also commonly referred to as ‘formulaic sequences’ (Wray, 2013). MWS are succinctly defined as variably-sized compositional recurring sequence patterns comprised of multiple words (for a recent overview, see Arnon and Christiansen, 2017). Three mechanisms that have been proposed to underpin frequency effects specifically in learning word sequences are described as follows (Diessel, 2007): [1] increased frequency causes the strengthening of linguistic representations, [2] increased frequency causes the strengthening of expectations and [3] increased frequency leads to the automatization of chunks. The frequency with which building blocks of language

occur is thus a driving force behind chunking and, all else being equal, each exposure to a given sequence of words (sounds or graphemes) will affect its subsequent processing. But why is there a need for chunking? To ameliorate the effects of the ‘real-time’ constraints on language processing imposed by the limitations of human sensory system and human memory in combination with the continual deluge of language input (cf., Christiansen and Chater, 2016a,b, for the ‘Now-or-Never bottleneck’), through constant exposure to (both auditory and visual) language input, humans learn to rapidly and efficiently recode incoming information into larger sequences. The fact that language is abundant in statistical regularities at multiple levels of language representations and that humans are able to detect such regularities via statistical learning allows for such chunking to take place. The by-products of statistical learning and chunking enable anticipatory language processing humans rely on to integrate the greatest possible amount of available information as fast as possible. Processing a MWS as a chunk will minimize memory load and speed up integration of the MWS with prior context (see, a chunk-based computational model presented in a recent study by McCauley and Christiansen, 2019).

1.2 MultiWord Frequency Effects in Online Processing

There is now an extensive body of evidence demonstrating that language users are sensitive to the input frequency across all levels of linguistic analysis (Ellis, 2002; Diessel, 2007; Jurafsky, 2003). An accumulating body of evidence now suggests that frequency effects also extend to the processing of MWS. Children and adults are shown to be sensitive to the statistics of MWS and rely on knowledge of such statistics to facilitate language processing and boost their acquisition (for overviews, see, Christiansen and Arnon, 2017; Shaoul and Westbury, 2011).

In the area of native language processing, a number of comprehension and production studies have provided evidence of processing advantages for MWS over non-MWS (see, e.g., Arnon and Snider, 2010; Bannard and Matthews, 2008; Conklin and Schmitt, 2012; Durrant and Doherty, 2010; Tremblay et al., 2011). Many of these studies follow an approach where the target stimuli are restricted to a certain frequency thresh-

old. The threshold-approach studies aimed to determine whether and to what extent MWS – i.e., more precisely ‘lexical bundles’ (LB)– are processed faster over less frequent counterparts (non-LB). The stimulus material is typically derived from language corpora based on predefined frequency criteria, while sequences differing in frequency matched on other properties were created as control stimuli. Biber and Conrad (1999) proposed that for a sequence of words to be considered to be considered a MWS, it must occur at least ten times per million in a corpus for sequences between two and four words long, and at least five times per million for longer sequences. Among these studies, (Tremblay et al., 2011) is the most relevant for the purposes of the present study. They created a dichotomous category for their stimuli based on Biber’s threshold criteria. Their sequences were matched on words in non-final position. Rather than presenting isolated phrases they embedded their sequences in the full sentential context, as in *I sat in the middle of the bullet train*. To examine sequence reading performance Tremblay et al. conducted three self-paced reading experiments: word-by-word reading, portion-by-portion reading and whole sentence reading. The three self-paced reading experiments showed that LBs have an online processing facilitatory effects over equivalent NLBs, i.e. in all of these experiments, sentences with LB were read faster than those with non-LB. The magnitude of the whole-string frequency effect increased with the length of the presentation window (i.e. word-by-word: $\sim 50 - 65ms$; portion-by-portion: $\sim 120ms$; sentence-by-sentence: $\sim 380ms$). The authors interpreted this incremental facilitatory effect as being linked to an increased opportunity to “skip” words.

While there has been an increased interest in the role of MWS in L2 online processing, most of the available research has focused on either non-compositional phrases, i.e. idioms (e.g. *kick the bucket*) or shorter compositional MWS including binomials (e.g. *bride and groom*) or collocations, i.e. frequently recurring two-word sequences (e.g. *perfectly natural*) (see, Conklin and Schmitt, 2012, for a review). However, much less is known whether and to what extent adult L2 speakers can develop sensitivities to the frequency of compositional – i.e. syntactically regular and semantically transparent – MWS larger

than two words. The few existing studies have produced inconsistent results: Some studies found frequency effects in processing of MWS in non-native speakers (e.g. Jiang and Nekrasova, 2007), whereas other studies found no such effects (e.g. Babaei et al., 2015). In addition, these previous studies have demonstrated frequency effects of MWS in a lexical (phrasal) decision task and/or acceptability judgment tasks using a self-paced reading paradigm.

1.3 The Present Study

As reviewed above, a processing advantage for MWS in native speakers is well attested. However, much less is known whether this extends to non-native (L2) speakers. The few existing L2 studies that have addressed this question have produced mixed results. The main goal of the present study is to replicate the processing advantage of MWS found for adult native language speakers based on evidence from self-paced reading and sentence recall tasks (Tremblay et al., 2011) in an ecologically more valid eye-tracking study in a group of L2 speakers. Eye movements of thirty participants were recorded using both early and late measures (first fixation duration, first-pass reading time, total reading time and fixation count). In line with emergentist accounts we predict that L2 speakers are sensitive to the statistics of MWS – to the frequencies of lexical bundles (LBs) – as evident in faster reading times across these four eye-tracking measures.

2 Method

2.1 Participants

Thirty L1 German L2 speakers of English (27 female) at the RWTH Aachen University participated in the study. There were 27 female and 3 male (mean age = 24.5; $SD = 5.1$). All participants had normal or corrected to normal vision. The L2 speakers were classified as having a Common European Framework (CEF) English proficiency level of upper intermediate (CEF = B2) or lower advanced (CEF = C1) based on their institutional status (educational background) and their scores on Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer and Broersma, 2012): an English vocabulary size test that is often used to estimate the CEF proficiency level. In addition, our participants completed the Language Experience and Proficiency Questionnaire

(LEAP-Q; Marian et al., 2007). Table 1 reports details on age of English acquisition, exposure, and proficiency of the L2 speakers group. The tested L2 group reached an average LexTALE score of 79.68, supporting their classification as intermediate to advanced. Regarding their English acquisition, the L2 speakers started learning English around the age of 9 and reported to have acquired fluency at around 15 years of age. On average, their current experience with English comes mainly from reading (mean score of 8.32 out of 10), watching TV (mean score of 7.61 out of 10) and listening to music (mean score of 6.55 out of 10). Self-ratings of their English language proficiency based on a 10-point scale were relatively high (all mean scores greater 7.5).

	mean (sd)	range
LexTALE score (average % correct)	79.68 (13.89)	55–100
<i>English acquisition</i> (years)		
Age start acquisition	9.29 (2.41)	1–13
Age became fluent	15.16 (3.64)	3–23
<i>Current experience</i>		
Family (0-10)	1.87 (2.68)	0–10
Friends (0-10)	3.74 (2.67)	0–10
Reading (0-10)	8.32 (2.12)	0–10
Music (0-10)	6.55 (3.24)	0–10
TV (0-10)	7.61 (2.22)	2–10
Self instruction (0-10)	4.16 (3.36)	0–9
Months in English-speaking country	2.89 (3.69)	0–14
<i>Self-rated</i> <i>L2 proficiency</i>		
Speaking (0-10)	7.55 (1.29)	5–10
Reading (0-10)	8.77 (1.06)	6–10
Listening (0-10)	8.52 (0.93)	6–10

Table 1: Summary of LexTALE scores and self-report information on English acquisition, exposure, and proficiency)

2.2 Material

We used the same stimulus material as in Tremblay et al. (2011). This material comprised of pairs of short sentences (mean length of sentences = 8.5 words ($SD = 0.7$)) that differed in exactly one word. An example of such a pair is presented in (1a) and (1b):

1a *I sat in the middle of the bullet train.*

1b *I sat in the front of the bullet train.*

The underlined portions in the sentences mark an MWS of either four or five words. The words in bold print are the words that distinguish MWS that are lexical bundles (LBs) – here *in the middle of the* – from those that are not (NLBs) – here *in the front of the*. Following Biber and Conrad (1999), the distinction of ‘lexical bundlehood’ was based on the frequencies of the MWS obtained from the spoken subcorpus of the BNC with frequency thresholds set to at least 10 occurrences per million words (for four-grams) and 5 occurrences per million (for five-grams). As shown in (1a) and (b), the MWS – LBs or NLBs – were embedded after the second word of the sentence and were followed by two more words. The frequency of the words occurring before and after the MWS were controlled. The sentence material comprised a total of 20 such pairs – 40 sentences containing LBs and NLBs – as well as 40 filler sentences (20 of which made sense and 20 were nonsensical). The sentence material was split into two counter-balanced lists, list A and list B, each of which contained 10 sentences that contained LBs, 10 sentences that contained NLBs, 10 filler sentences that were meaningful, and 10 filler sentences that were nonsensical. A complete list of the stimulus material can be found in Tremblay et al. (2011).

2.3 Procedure

Participants were randomly assigned to one of two groups. Group one was first presented list A, followed by a thirty minute break, followed by list B. Group 2 was presented with the two lists in reversed order. The sentences were presented on a 23-inch TFT monitor (resolution: 1920 x 1080 pixels) in pseudorandomised order, i.e. order of presentation was randomly determined but then kept constant across groups. Participants were instructed to read the sentences for comprehension silently and at their own pace. Each trial consisted of the following steps. The participants saw an asterisk in the center of the screen (font: Arial bold; size: 100). When ready, the participants pressed a key to see the first sentence, which was then displayed in a single line with black 30-point font characters on a white background at the centre. Once they had finished reading the sentence, participants pressed a key to see the next one. Each trial ended with a simple yes-no question specific to the sentence to ensure that the participants ac-

tually read and processed the material. Eye movements were recorded using a Tobii Tx300 remote eye tracker that records binocular gaze data at 300 Hz and filtered with the Tobii fixation filter with standard settings (velocity threshold = 30 pixels/sample; distance threshold). The experiment took about 15 minutes (incl. calibration and explanation).

2.4 Statistical analysis

Eye movements were analyzed based on data collected from four measures: (1) first fixation duration (FFD), i.e. time spent initially fixating the MWS region, (2) first pass reading time (FPRT), i.e. sum of all the fixations made in the MWS region until the point of fixation leaves the region, (3) total reading time (TRT), i.e. sums all fixation times made within a MWS region, including those fixations made when re-reading the region and (4) the number of regressive saccades into the MWS region (COUNT).¹ L2 speakers' sensitivity to MWS frequency was evaluated using mixed-effect regression models implemented with the lme4 package (Bates et al., 2014) in the R environment (R Core Team, 2018). Separate models were fitted for each of the four dependent measures gathered in the eye-tracking experiment (FFD, FPRT, TRT, COUNT). Fixation times were logged (natural log) to reduce the nonnormality of their distributions. In each model, the dependent measure was regressed onto the predictor lexical bundlehood (dummy coded: LB vs. NLB). In addition, two control variables (length of MWS (in characters) and participants' LexTALE scores, a measure of L2 vocabulary size) were entered into each model as fixed effects. All models had the maximal random-effects structure justified by the design (Barr et al., 2013), which included by-subject random intercepts and slopes for lexical bundlehood as well as random intercepts for items.

3 Results

Prior to the analyses, – for each eye tracking measure – all trials that were more than 2 standard deviations above or below the participant's mean

¹FFD and FPRT are 'early measures' that are indicative of early processes during reading (e.g. familiarity checks, access to orthographic/phonological information and lexical meaning, cf. Reichle et al., 1998). TRT and COUNT are 'late measures' taken to reflect later processes (e.g. reanalysis of information, integration of information in discourse and recovery from processing difficulties; cf. Rayner, 1998).

score were removed. This led to a loss of data of about 5% (4.9% for FFD, 4.4% for FPRT, 4.7% for TRT, and 4.8% for COUNT). Figure 1 shows the distributions of all four eye-tracking dependent measures for multiword sequences (MWS) that are lexical bundles (LB; left) and those that are not lexical bundles (NLB; right). The plots in Figure 1 suggest a processing advantage of MWS that are LB over those that are NLB for three out of the four eye tracking measures. On average, participants exhibited shorter total reading times (TRT) (LB: $M = 1397.59$, $SD = 672.38$, $Median = 1283.00$; NLB: $M = 1752.64$, $SD = 730.06$, $Median = 1681.50$), shorter first pass reading times (FPRT) (LB: $M = 800.75$, $SD = 374.76$, $Median = 766$; NLB: $M = 1002.61$, $SD = 490.85$, $Median = 943$) and a smaller number of regressive saccades to the region of interest (COUNT) (LB: $M = 6.46$, $SD = 3.04$, $Median = 6$; NLB: $M = 7.82$, $SD = 3.36$, $Median = 8$). The duration of first fixation (FFD) was about the same across LBs and NLBs (LB: $M = 222.44$, $SD = 63.04$, $Median = 219$; NLB: $M = 232.38$, $SD = 62.25$, $Median = 221.92$).

The results of the mixed effects models are presented in Table 2. The top part of Table 2 presents the information regarding the effects of our key predictor variable, lexical bundlehood, and the two control variables, MWS length (in characters) and LexTALE scores. The 'Intercept' row lists the mean fixation times (for the TRT, FPRT and FFD measures) and regressive saccade count (for the COUNT measure) for LBs on the log scale. The 'NLB' row indicates the difference in log fixation times – or, in the case of the COUNT model, regressive saccade counts – between LBs and NLBs. The results show that – for all dependent variables except FFD, which only approached significance – lexical bundlehood was found to be a significant predictor of eye movements, even after controlling for the effects of MWS length and LexTALE scores: Participants were significantly faster in processing sentences containing LBs compared to sentences containing equivalent control items with NLBs $NLB_{TRT}: estimate = 0.33$, $SE = 0.09$, $t = 3.9$, $p < 0.001$). After accounting for the effects of length and L2 proficiency and adjusting for the individual variation between subjects and items, sentences with LBs were read ($exp(6.76 + 0.33) - exp(6.76) =$) 339

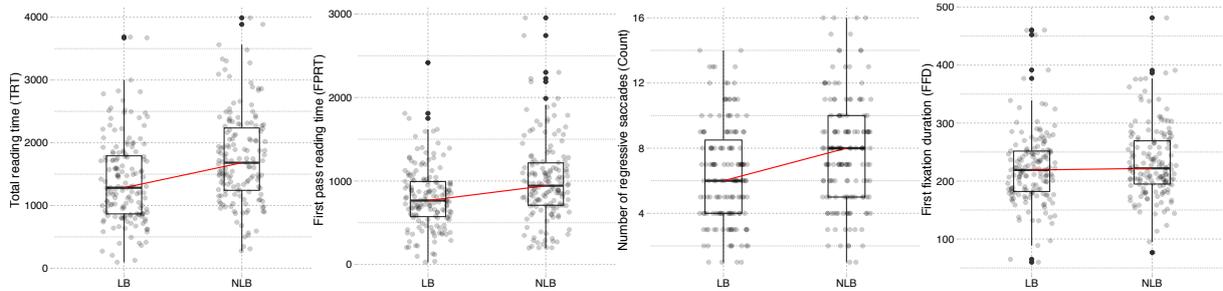


Figure 1: Distributions of all four dependent measures from the eye-tracking experiments for multiwords sequences that are lexical bundles (LB) or not (NLB). Red lines connect the median values of the two contrasted MWS types.

ms faster on average. In comparison, (Tremblay et al., 2011) report a processing advantage of LB of 380ms in their self-paced reading time study (sentence-by-sentence presentation) on native English speakers. First pass reading times (FPRT) of LBs were ($\exp(6.13 + 0.33) - \exp(6.13) = 181$ ms faster (NLB_{FPRT} : $estimate = 0.33$, $SE = 0.08$, $t = 4.41$, $p < .001$). Regarding the COUNT measure, the model predicted an average increase of ($\exp(1.544 + 0.281) - \exp(1.544) = 1.52$ regressive saccades for NLBs relative to LBs (NLB_{COUNT} : $estimate = 0.28$, $SE = 0.08$, $t = 3.58$, $p < .001$). The difference in first fixation duration between LBs and NLB was just over ($\exp(5.316 + 0.056) - \exp(5.316) = 11$ ms, which was marginally significant (NLB_{FFD} : $estimate = 0.06$, $SE = 0.01$, $t = 1.81$, $p = 0.07$). The bottom part of Table 2 presents the variability in the data that is attributable to random effects (e.g. some participants exhibited overall faster reading times than others). We found that – across the four eye tracking measures - there was a relatively large amount of variability in reading speed between participants (TRT: $SD = 0.47$, FPRT: $SD = 0.39$; FFD: $SD = 0.09$; COUNT: 0.42) and relatively little between items (TRT: $SD = 0.16$, FPRT: $SD = 0.09$; FFD: $SD = 0$; COUNT: 0.15). The standard deviation for the by-subject random slopes for lexical bundlehood were minimal (all $SD < 0.2$), indicating that the LB effect was consistent across subjects. This pattern of results is line with the results reported in (Tremblay et al., 2011).

4 Discussion

The main goal of the present study was to determine whether non-native (L2) speakers can develop sensitivity to the statistics of composi-

tional multiword sequences (MWS) larger than two words. To this end, the study aimed to replicate the processing advantage of such sequences found for native speakers (Tremblay et al., 2011) in a group of L2 speakers of English. As reviewed in Section 1, (Tremblay et al., 2011) performed three self-paced reading studies to investigate the facilitatory effects of lexical bundles (LBs) and found that the magnitude of the whole-string frequency effect increased with the length of the presentation window. We were able to replicate the MWS frequency effects using eye-tracking methodology: Mixed-effects modeling revealed that lexical bundlehood was a significant predictor of eye movements, even after controlling for the effects of MWS length and LexTale scores and after adjusting for the individual variation between subjects and items: Participants were significantly faster in processing sentences containing LBs compared to sentences containing equivalent control items with NLBs for all dependent variables except first fixation duration (FFD), which approached significance ($p < 0.1$). Similar results were reported in a recent eye-tracking study on the online processing of multiword sequences in Chinese (see, Yi et al., 2017) where significant or marginally significant effects of MWS frequency were found in the eye movement measures also investigated in the present study. Like the present study, (Yi et al., 2017) found that the effect of FFD on reading times was marginally significant. The findings reported here are thus consistent with the results reported in previous L2 studies (Ellis, 2008; Durrant and Schmitt, 2009; Hernández et al., 2016; Jiang and Nekrasova, 2007; Kerz and Wiechmann, 2017; Siyanova-Chanturia et al., 2011). Our study thus provides additional evidence in support of the hypothesis that similarly to native speakers, non-native speakers can also

	<i>Dependent variable:</i>			
	log(TRT)	log(FPRT)	log(FFD)	COUNT
	<i>Linear</i>	<i>Linear</i>	<i>Linear</i>	<i>Poisson</i>
	<i>mixed-effects</i>	<i>mixed-effects</i>	<i>mixed-effects</i>	<i>mixed-effects</i>
	(1)	(2)	(3)	(4)
<i>Fixed effects:</i>				
(Intercept)	6.766*** (0.435)	6.133*** (0.370)	5.316*** (0.162)	1.544*** (0.491)
NLB	0.331*** (0.085)	0.332*** (0.075)	0.056 (0.031)	0.281*** (0.079)
MWS Length	0.049*** (0.014)	0.056*** (0.011)	0.003 (0.005)	0.055*** (0.013)
LexTALE	-0.004 (0.005)	-0.006 (0.004)	-0.00000 (0.002)	-0.006 (0.006)
<i>Random effects:</i>				
Std.Dev Subject (Intercept)	0.351	0.395	0.091	0.418
Std.Dev Subject LB	0.046	0.177	0.031	0.087
Std.Dev Item (Intercept)	0.159	0.093	0.000	0.147
Residual	0.434	0.478	0.269	—
<i>Note:</i>	* $p < .05$; ** $p < .01$; *** $p < .001$			

Table 2: Regression coefficients (with standard errors) from the four mixed-effects models fitted to the eye-movement data. Estimates and standard errors of fixation times are in logged milliseconds. One observation is equal to one fixation time (or – in the case of the COUNT-model – regressive saccade count) measurement for one sentence read by one participant.

develop the sensitivity to the statistics of MWS. At a more general theoretical level, the results of the present study are consistent with emergentist accounts that challenge dual-system views of language and instead argue for single-systems of language. More importantly, the results indicate that similarities between L1 and late L2 learning are more striking than the differences and, therefore, that unified theoretical models rather than separate ones are needed to account for the mechanisms used for L1 and L2 learning (see, e.g., [MacWhinney, 2017](#)). Emergentist accounts have proposed such mechanisms, namely that of statistical learning and chunking (see Section 1.1 for more details). Sensitivity to the statistics of multiword sequences facilitates chunking - required to integrate the greatest possible amount of available information as fast as possible so as to overcome the fleeting nature of linguistic input and the limited nature of our memory for sequences of linguistic input (Now-or-Never bottleneck, see [Christiansen and Chater, 2016a](#)).

Some of the questions left open by the current study may provide interesting avenues for future work. First, we investigated sensitivity to ‘simple statistics’ – i.e. corpus-derived frequencies – of MWS in non-native speakers. The question arises whether similar results could be obtained for ‘more complex’ distributional statistics using association measures, such as transitional probability or mutual information or using information-theoretic measures, such as entropy as well as measures that capture the variability of MWS. Second, the stimulus material used in this study was derived from a corpus representing spoken language. In the light of growing evidence that the statistics of written input play a crucial role in the development of linguistic knowledge – as it provides a source of substantial change in the statistics of an individual’s language experience ([Seidenberg and MacDonald, 2018](#)) – it would be important to determine whether language users can ‘tune to’ multiple statistics inherent in different registers/genres. And, third, it would be important to determine whether the ability to tune to the statistics of MWS is subject to individual differences, and if so, to what extent these differences are linked to a host of experience-related, cognitive and affective factors.

References

- Inbal Arnon and Morten H Christiansen. 2017. The role of multiword building blocks in explaining 11–12 differences. *Topics in Cognitive Science*, 9(3):621–636.
- Inbal Arnon and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.
- Samaneh Babaei, Ferdos Taleb Najafabadi, and Zahra Fotovatnia. 2015. Processing of lexical bundles by persian speaking learners of english. *Journal of Teaching Language Skills*, 33(4):1–18.
- Colin Bannard and Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children’s repetition of four-word combinations. *Psychological Science*, 19(3):241–248.
- Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.
- Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, et al. 2014. lme4: Linear mixed-effects models using eigen and s4. *R package version*, 1(7):1–23.
- Clay Beckner, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1):1–26.
- Douglas Biber and Susan Conrad. 1999. Lexical bundles in conversation and academic prose. *Language and Computers*, 26:181–190.
- Franklin Chang, Gary S Dell, and Kathryn Bock. 2006. Becoming syntactic. *Psychological Review*, 113(2):234.
- Nick Chater and Morten H Christiansen. 2018. Language acquisition as skill learning. *Current opinion in behavioral sciences*, 21:205–208.
- Morten H Christiansen and Inbal Arnon. 2017. More than words: The role of multiword sequences in language learning and use. *Topics in cognitive science*, 9(3):542–551.
- Morten H Christiansen and Nick Chater. 2016a. *Creating language: Integrating evolution, acquisition, and processing*. MIT Press.
- Morten H Christiansen and Nick Chater. 2016b. The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39.
- Kathy Conklin and Norbert Schmitt. 2012. The processing of formulaic language. *Annual Review of Applied Linguistics*, 32:45–61.

- Christopher M Conway, Althea Bauernschmidt, Sean S Huang, and David B Pisoni. 2010. Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3):356–371.
- Holger Diessel. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, 25(2):108–127.
- Philip Durrant and Alice Doherty. 2010. Are high-frequency collocations psychologically real? investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2):125–155.
- Philip Durrant and Norbert Schmitt. 2009. To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, 47(2):157–177.
- Nick Ellis. 2008. The associative learning of constructions, learned attention, and the limited L2 endstate. In Peter Robinson and Nick Ellis, editors, *Handbook of cognitive linguistics and second language acquisition*, chapter 15, pages 372–405. Routledge.
- Nick C Ellis. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2):143–188.
- Nick C. Ellis. 2019. Essentials of a theory of language cognition. *The Modern Language Journal*, 103(S1):39–60.
- Nick C Ellis and Diane Larsen-Freeman. 2006. Language emergence: Implications for Applied Linguistics. Introduction to the special issue. *Applied Linguistics*, 27(4):558–589.
- Marc Ettliger, Kara Morgan-Short, Mandy Faretta-Stutenberg, and Patrick Wong. 2016. The relationship between artificial and second language learning. *Cognitive Science*, 40(4):822–847.
- Ram Frost, Noam Siegelman, Alona Narkiss, and Liron Afek. 2013. What predicts successful literacy acquisition in a second language? *Psychological Science*, 24(7):1243–1252.
- Mireia Hernández, Albert Costa, and Inbal Arnon. 2016. More than words: multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, 31(6):785–800.
- Ray Jackendoff and Ray S Jackendoff. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, USA.
- Nan AN Jiang and Tatiana M Nekrasova. 2007. The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3):433–445.
- Dan Jurafsky. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In Rens Bod, Jennifer Hay, and Stefanie Jannedy, editors, *Probabilistic linguistics*, volume 21. Mit Press Cambridge, MA.
- Elma Kerz and Daniel Wiechmann. 2017. Individual differences in L2 processing of multi-word phrases: Effects of working memory and personality. In *Computational and Corpus-Based Phraseology. EUROPHRAS 2017*, pages 306–321. Lecture Notes in Computer Science, Springer. vol 10596.
- Evan Kidd and Joanne Arciuli. 2016. Individual differences in statistical learning predict children’s comprehension of syntax. *Child development*, 87(1):184–193.
- Evan Kidd, Seamus Donnelly, and Morten H Christiansen. 2017. Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, pages 154–169.
- Jill Lany and Jenny R Saffran. 2010. From statistics to meaning: Infants’ acquisition of lexical categories. *Psychological Science*, 21(2):284–291.
- Kristin Lemhöfer and Mirjam Broersma. 2012. Introducing lextale: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, 44(2):325–343.
- Willem JM Levelt. 1993. *Speaking: From intention to articulation*, volume 1. MIT press.
- Brian MacWhinney. 2012. The logic of the unified model. In Susan M. Gass and Alison Mackey, editors, *The Routledge handbook of second language acquisition*, pages 211–227. Routledge London & New York.
- Brian MacWhinney. 2017. A shared platform for studying second language acquisition. *Language Learning*, 67(S1):254–275.
- Brian MacWhinney and William O’Grady. 2015. *The handbook of language emergence*. John Wiley & Sons.
- Viorica Marian, Henrike K Blumenfeld, and Margarita Kaushanskaya. 2007. The language experience and proficiency questionnaire (leap-q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*.
- Jessica Maye, Daniel J Weiss, and Richard N Aslin. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1):122–134.
- Stewart M McCauley and Morten H Christiansen. 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological review*, 126(1):1.
- James L McClelland, Matthew M Botvinick, David C Noelle, David C Plaut, Timothy T Rogers, Mark S Seidenberg, and Linda B Smith. 2010. Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences*, 14(8):348–356.

- Jennifer B Misyak and Morten H Christiansen. 2012. Statistical learning and language: An individual differences study. *Language Learning*, 62(1):302–331.
- Rosamond Mitchell, Florence Myles, and Emma Josephine Marsden. 2013. *Second Language Learning Theories: Third edition*, 3rd edition. Routledge.
- Luca Onnis, Stefan L Frank, Hongoak Yun, and Matthew Lou-Magnuson. 2016. Statistical learning bias predicts second-language reading efficiency. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, pages 2105–2110.
- Luca Onnis, Heidi R Waterfall, and Shimon Edelman. 2008. Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3):423–430.
- Sébastien Pacton, Michel Fayol, and Pierre Perruchet. 2005. Children’s implicit learning of graphotactic and morphological regularities. *Child Development*, 76(2):324–339.
- Steven Pinker. 1999. *Words and rules: The ingredients of language*. Basic Books.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.
- Martin Redington and Nick Chater. 1997. Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences*, 1(7):273–281.
- Erik D Reichle, Alexander Pollatsek, Donald L Fisher, and Keith Rayner. 1998. Toward a model of eye movement control in reading. *Psychological review*, 105(1):125.
- Jenny R Saffran, Richard N Aslin, and Elissa L Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928.
- Jenny R Saffran and Diana P Wilson. 2003. From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy*, 4(2):273–284.
- Mark S Seidenberg and Maryellen C MacDonald. 2018. The impact of language experience on language and reading: A statistical learning approach. *Topics in Language Disorders*, 38(1):66–83.
- Cyrus Shaoul and Chris Westbury. 2011. Formulaic sequences: Do they exist and do they matter? *The mental lexicon*, 6(1):171–196.
- Noam Siegelman and Ram Frost. 2015. Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81:105–120.
- Anna Siyanova-Chanturia, Kathy Conklin, and Norbert Schmitt. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2):251–272.
- Erik D Thiessen and Jenny R Saffran. 2003. When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, 39(4):706.
- Susan P Thompson and Elissa L Newport. 2007. Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3(1):1–42.
- Antoine Tremblay, Bruce Derwing, Gary Libben, and Chris Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2):569–613.
- Alison Wray. 2013. Formulaic language. *Language Teaching*, 46(3):316–334.
- Wei Yi, Shiyi Lu, and Guojie Ma. 2017. Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, 33(4):519–549.