Title

Tuning to Multiple Statistics:Second Language Processing of Multiword Sequences across Registers

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Kerz, Elma
Wiechmann, Daniel
Christiansen, Morten H.

Publication Date

2019

Peer reviewed

Tuning to Multiple Statistics: Second Language Processing of Multiword Sequences across Registers

Elma Kerz (elma.kerz@ifaar.rwth-aachen.de)
Department of English Linguistics, RWTH Aachen University, Karmanstr. 17-19
52064 Aachen, Germany

Daniel Wiechmann (d.wiechmann@uva.nl)
Institute for Logic, Language and Computation, University of Amsterdam, Spuistraat 134
Amsterdam, 1012 VT Netherlands

Morten H. Christiansen (christiansen@cornell.edu)
Department of Psychology, Cornell University, Uris Hall
Ithaca, New York 14853

Abstract

A substantial body of research has demonstrated that children and adults (both native and non-native speakers) are sensitive to the statistics of multiword sequences (MWS) and rely on knowledge of such statistics to facilitate their language processing and boost their acquisition. However, this research was primarily aimed at determining whether and to what extent speakers can develop sensitivity to MWS statistics of a single type of linguistic input: that of spoken language. Recently, there has been a growing awareness of the key role of written input in the development of linguistic knowledge, as it provides a source of substantial change in the statistics of an individual's language experience. The present study reports on a series of experiments designed to determine whether second language learners of English are able to develop sensitivity to distributional statistics of MWS inherent in different (register-specific) input types.

Keywords: life-long learning; multiword sequences; second language processing; statistical learning

Recent theoretical approaches have highlighted the key importance of linguistic experience to the acquisition and processing of language. This broad class of approaches, commonly referred to as 'emergentist' approaches,[1] 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'). Language acquisition is viewed as learning how to process linguistic input efficiently (e.g., Chang, Dell & Bock, 2006; Chater & Christiansen, 2018). In the emergentist perspective, language learning does not result in the establishment of a static knowledge system. Rather, as long as there is exposure to linguistic input, an individual's knowledge of a language is subject to constant change. Learning about the statistical regularities and distributional patterns inherent in linguistic input is viewed as a continuous process that does not end at some discrete point in time in ontogenetic development but instead takes place across the lifespan (e.g., Armstrong et al., 2017; Saffran & Kirkham, 2018; Seidenberg & MacDonald, 2018). This lifelong process brings about changes in language representations in response to the statistics in linguistic input. These experientially-driven adaptive processes are shown to occur across multiple linguistic levels and apply to the acquisition of new structures, the modification and/or adjustment of already learned representations or changes in accessibility of learned representations.

Moving away from the traditional 'words-and-rules' approach (e.g., Pinker, 1999), emergentist accounts have developed an increasing interest in the role of multiword sequences (MWS), often defined as variably-sized continuous or discontinuous recurring strings of words. This interest stems from an extensive body of evidence demonstrating that children and adults (both native and non-native speakers) are sensitive to the statistics of MWS and rely on knowledge of such statistics to facilitate their language processing and boost their acquisition (e.g. Shaoul & Westbury, 2011; N. Ellis, 2011; see Arnon & Christiansen, 2017, for a recent review). Sensitivity to the statistics of MWS facilitates chunking - required to integrate the greatest possible amount of available information as fast as possible so at 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 & Chater, 2016). Processing a MWS as a chunk will minimize memory load and speed up integration of the MWS with prior context (see, e.g., a chunk-based computational model of early language acquisition presented in a recent study by McCauley & Christiansen, 2019).

Frequency estimates obtained from corpora of actual language use have been shown to be robust predictors of language behavior across different types of experimental designs, as evidenced by higher accuracy rates, faster reaction times, and fewer and faster fixations. These effects have been shown in both child and adult populations as well as second-language learner populations. While earlier studies on the processing of MWS have used a threshold-approach to test whether MWS are stored and processed as holistic units (Biber & Conrad, 1999), more recent studies have incorpo-

[1] Following the literature (see, e.g., Kidd et al. 2018), we use the term 'emergentist' to refer to a broad class of approaches - usage-based (a.k.a. experience-based) models, complex dynamic systems theory, constraint-based approaches, exemplar-based models and connectionist models - that share a number of key tenets, for more details (see, e.g., Beckner et al. 2009; Daelemans & van den Bosch, 2005; McClelland et al. 2010)

rated further methodological improvement by testing these effects across the frequency continuum after controlling for substring frequency (for studies in child language acquisition, see, Bannard and Matthews, 2008; Matthews and Bannard, 2010; for studies on adult – both first and second language – processing see, Arnon, McCauley & Christiansen, 2017; Arnon and Snider, 2010; Hernandez et al., 2016, Kerz & Wiechmann, 2017, Yi et al. 2017).

This line of research has also shown that while being the most robust statistic, frequency is not the only kind of distributional information to which language users are sensitive. For example, in a study of MWS repetition in children, Matthews and Bannard (2010) showed that MWS with a high slot entropy value have increased uncertainty for what word occur in that slot and that such sequences were easier to generalize and hence easier to process for children than MWS with lower slot entropy.

The prior studies reviewed here have made important theoretical and methodological contributions to research on MWS. However, they have primarily focused on examining sensitivity to the frequencies derived from corpora representing spoken language (i.e., spontaneous conversations). In contrast to early child language acquisition (prior to literacy), where children are mainly exposed to the statistics of the spoken linguistic input (i.e., to child-directed speech), the role of written language becomes increasingly more important in later stages of learning which also sees increased demands on literacy. Indirect support for this assumption comes from a growing number of studies indicating that written language constitutes a key input type in the development of linguistic knowledge, as it provides a source of substantial change in the statistics of an individual's language experience (e.g., Montag & MacDonald, 2015; Seidenberg & MacDonald, 2018). Language users are thus faced with the challenge of keeping track of the ever-changing statistics of these two main types of language input. This challenge is exacerbated by considerable variability in the distributional properties of linguistic patterns at multiple levels of linguistic structure *within* these two input types (Roland, Dick & Elman, 2007; see also work on register/genre[2] variation by Biber and colleagues, e.g. Biber et al. 1999, Biber & Conrad, 2009).

In light of the lifelong nature of language learning highlighted in emergentist accounts, there is an apparent need not only to characterize the statistical learning processes in early stages of child language development, but also to understand how language users develop sensitivity in later stages of learning to the multiple kinds of statistics found in written language. This issue is of particular importance for second language (L2) learners, who are likely to get a lot of their language from written sources. Using a within-subject design, the present study sets out to investigate whether and to what extent language users can develop sensitivity towards the multiple statistics of MWS. We perform analyses of large samples of corpus data representing four registers and use the results from these analyses to make predictions about language users' performance in a MWS decision task. We predict faster response latencies for more frequent MWS (after controlling for all part frequencies) across the registers/genres investigated here. In addition to determining the effects of frequency ('more simple' distributional statistics), the study also investigated whether and to what extent language users are sensitive to 'more complex' distributional statistics (entropy) that captures the variability of MWS. The effects of frequency and entropy were investigated in a L2 learner population by conducting four reaction time experiments where processing latencies of MWS are compared for pairs of MWS that differ in sequence-frequency and entropy of their final slot.

[2]In the present paper, the terms 'registers' and 'genres' are used interchangeably in Biber's sense (2006:11) as referring to "situationally-defined varieties described for their characteristic distributions of linguistic structures and patterns."

Methods

Participants

Sixty advanced learners of English participated as a part of a larger project (34 female and 26 male, M = 23.56 years, SD = 4.52). All participants were college students recruited from the RWTH Aachen University studying either towards an BA or an MA at the time of testing. Participants were asked to fill out the Language Experience and Proficiency Questionnaire (LEAP-Q, see, Lemhofer & Broersma, 2012), a questionnaire used to obtain general demographic information and more specific information on self-rated proficiency for three language areas (reading, understanding and speaking) and self-rated current knowledge of L2 English and exposure to the L2. The data gathered from the LEAP-Q instrument are reported in Table 1, showing means, standard deviations and ranges of our L2 group.

Materials

The current study follows the general methodological approach described in the previous studies reviewed above that used carefully chosen stimuli, controlled for substring frequency. Following these studies, we chose pairs of four-word sequences as stimuli that differed only in the final word and in overall MWS frequency (high vs. low) but were matched for substring frequency (e.g. *to justify the cost* vs. *to justify the effort* from the newspaper register; e.g., *is beyond the scope* vs. *is beyond the boundaries* from the academic register). We constructed a total of 240 experimental items, 60 for each of four registers. The items were constructed using the Corpus Contemporary American English (COCA; Davies, 2008), a 560 million words corpus with approximately equal-sized subcomponents representing the statistics of MWS from the four target registers: (1) spoken (118 million words), (2) fiction: (113 million words), (3) newspaper (114 million words and (4) academic journals (112 million words). In a first step, all COCA text files were preprocessed using the sentence splitting (`ssplit`) and tokenization (`PTNTokenizer`) components from the Stanford CoreNLP toolkit V.3.2.9 (Manning et

Table 1: Self-report information on English acquisition, exposure, and proficiency

	mean	sd	obs. range
English acquisition (years)			
Age start acquisition	8.46	2.23	6–22
Age became fluent	14.63	3.9	8–23
Current exposure to English			
Friends (0-10)	4.63	3.1	0–10
Family (0-10)	1.36	2.6	0–10
Reading (0-10)	7.64	2.25	1–10
Class instruction (0-10)	5.48	3.43	0–10
Self instruction (0-10)	4.86	2.81	0–10
Watch (0-10)	7.64	2.72	0–10
Listening music (0-10)	7.39	2.84	0–10
Social Media (0-10)	7.39	2.68	0–10
Immersion (month)			
English speak. country	2.96	3.46	0–11
Self-rated proficiency			
Speaking (0-10)	7.25	1.69	1–10
Listening (0-10)	8.49	1.38	5–10
Reading (0-10)	7.86	1.58	1–10

al., 2014). In a second step, we extracted frequencies for all n-grams of orders 1 to 4 using Java scripts. N-grams with a frequency of one (so-called 'hapax legomena') were discarded. These two steps were performed on the RWTH Aachen University high-performance computing cluster. In a third step, four-grams that had a function word as their last word were filtered out to ensure that the position at which entropy was measured was filled by a lexical word.[3]. For all remaining four-grams the Shannon entropy H was computed for their final word slot, which is given in (1), where X is the final slot of the MWS, each x is a word that appears in that slot, and $p(x)$ is the probability of seeing each x in that position. All conditional word probabilities needed to compute entropy scores were estimated using second-order Markov models (cf. Willems, Frank, Nijhof, Hagoort, and van den Bosch, 2015).

$$H(X) = -\sum_{x} p(x) \log p(x) \quad (1)$$

We next identified all sequences of four words that began with the same first three words (i.e. shared the same pattern). Within each set of these sequences, a frequency difference score (FDS) was computed for a given sequence in relation to the most frequent sequence in that set.[4] We then ordered the sequences according to their FDS and explored how FDS scores related to entropy using a moving window approach/technique. A window with a size that corresponded to a predefined FDS was moved over the entire candidate-item pool to bin all four-gram into groups with similar FDS (see Figure 2 for a visualization). Inspection of these data indicated that four-grams with small differences in FDS tended to exhibit low entropy scores. Based on these observations, we restricted our candidate pool to four-grams that had entropy scores between 0 and 3 and a difference in log normalized frequency between 6.5 and 30. From this candidate pool, we randomly sampled, from each register, a total of 60 experimental item pairs: 20 pairs from each of three entropy ranges ('low': H(X) between 0 and 1, 'mid': H(X) between 1 and 2, 'high': H(X) between 1 and 2). Applying these filters meant that the log frequencies of our items ranged between 0.69 and 6.85 (spoken $0.69-6.07$, fiction: $0.69-6.85$; news $0.69-5.97$, academic $0.69-5.87$).

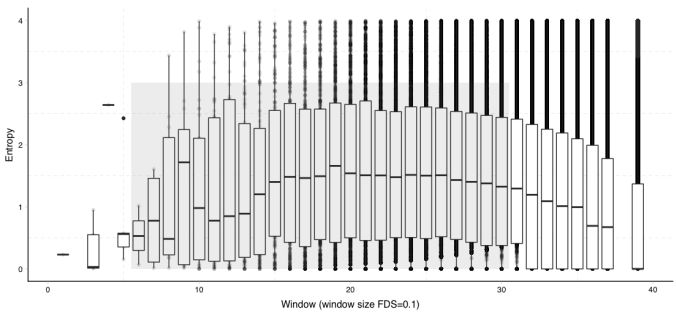

Figure 1: Distribution of entropy scores across the 'frequency difference score' (FDS) range for the academic subcomponent of COCA. The shaded area represents the range from which experimental items were sampled.

Procedure

Four separate MWS experiments, one for each register, were conducted as part of a larger project at two different days with two registers tested per day (day 1: academic and fiction; day 2: news and spoken). Each experiment was divided into two blocks of about 7 minutes each, which were separated by intervening tasks assessing individual differences in L2 experience and another task assessing a cognitive individual differences variable (not investigated here). The 120 MWS from a given register were distributed across two lists that each contained one of the two variants of a given pair, so that in a given experimental run participants would never see both variants of a pair. In addition to the experimental items, the lists also contained 60 ungrammatical items, which were incorrect due to scrambled word order. The order of presentation of the blocks was counterbalanced between participants. Participants were asked to judge if a four-word sequence that appeared on the screen was a possible sequence in English or not. They were given no information about the fact that MWS were extracted from different registers. Each trial began with the presentation of a fixation point for 500 ms. Phrases appeared at once in the middle of the screen and participants were instructed to respond as quickly and accurately as possible using the keyboard. The MWS was then presented and stayed visible on the screen until participants responded or

[3]The stop-list for function words was derived from the 'Essential Word List' https://www.edu.uwo.ca/faculty-profiles/docs/other/webb/essential-word-list.pdf

[4]FDS were expressed in terms of as the absolute of the $log10$ of the normalized frequency of a four-gram minus the $log10$ of the normalized frequency of the most frequent four-gram sharing the first trigram.

until 3000 ms had passed. The task was run using PsychoPy v3.0 (Peirce, 2007).

Results

Responses under 200 ms and over two standard deviations from the mean were excluded. This resulted in loss of small percentage of data for each register ($< 9\%$). Accuracy for target items was near ceiling ($> 92\%$ correct) for all registers. On average, participants were faster in responding to MWS from the spoken and fiction registers (mean response latencies spoken and fiction = 1.44 seconds, $SD = 0.5$) than in responding to MWS from the academic and news registers (mean response latencies: academic = 1.55 seconds, $SD = 0.51$; news = 1.57 seconds, $SD = 0.5$). The results were analyzed using mixed-effect linear regression models. To determine to what extent L2 learners can develop sensitivity to the two distributional statistics of MWS (log MWS frequency, slot entropy) inherent in the four registers investigated in the study, separate models were fitted to the data from each of the four experiments.[5] All analyses were carried out using the `lme4` package (v 1.1-17, Bates et al., 2015) in `R` (version 3.5.0; R Core Team, 2017). Log response times were used as the predicted variable to reduce the skewness in the distribution of response times. In a first step, we fitted models containing all control variables, i.e. LENGTH (in number of characters), two substring frequency measures (LOG FREQUENCY OF THE FINAL UNIGRAM and LOG FREQUENCY OF THE FINAL BIGRAM)[6], BLOCK ORDER (first vs. second), and PAIR VARIANT (high-low frequency variant of a pair). All continuous predictors were mean centered prior to analysis. We then added our key predictors, LOG PHRASE FREQUENCY and SLOT ENTROPY (high, mid, low), to examine their predictive value over and above our controls, using likelihood ratio tests to compare nested models. In all models, we used the maximal random effect structure justified by the data, which included random intercepts for participants and items and by-subject random slopes for log MWS frequency and entropy. To compare the effects of (log) MWS frequency on (log) reaction times across the four registers, standardized coefficients as well as marginal and conditional pseudo-R2 were computed (cf. Nakagawa and Schielzeth, 2013).[7]

In a next step, we tested for a potential interaction between our two key predictors. To this end, we conducted model comparisons between a model containing only the main effects and a corresponding model that also included the two-way interaction between MWS frequency and slot-entropy using Akaike's Information Criterion (AIC). The results of the final best-fitting model for each register are presented in Table 2 below. Likelihood ratio tests comparing models including LOG MWS FREQUENCY with a model that included only the control variables revealed that – after statistically controlling for the effects of length and frequency-related control variables – MWS frequency exerted a significant effect on (log) reaction times for all registers except fiction (spoken: $\chi(1) = 18.53$, $p < .0001$; fiction: $\chi(1) = 2.28$, $p = 0.51$; academic: $\chi(1) = 13.31$, $p = 0.004$; news: $\chi(1) = 59.46$, $p < .0001$). The frequency effect was found to be strongest in the spoken and news registers (both standardized $\beta = -0.13$), followed by academic language (standardized $\beta = -0.11$). A significant main effect of slot entropy was observed for the spoken and academic register (spoken: $\chi(1) = 40.03$, $p < 0.001$; fiction: $\chi(1) = 5.77$, $p = 0.58$; academic: $\chi(1) = 14.23$, $p = 0.047$; news: $\chi(1) = 12.42$, $p = 0.061$), such that that mean response times were significantly faster for MWS with higher slot entropy. There was also a significant interaction effect between MWS frequency and entropy in the spoken register ($\beta = -0.046$, $SE = 0.016$, $t = -2.79$), indicating that the frequency effect was more pronounced in high-entropy MWS (see Figure 2). The effects of the length and frequency related control variables were in the predicted directions - with longer MWS being read more slowly on average and more frequent final words leading to faster response times - but these effects were significant in only some of the registers. Significant effects of block order were observed for two of the four registers (see Table 3 for details).

Discussion and Conclusions

In this paper we reported a series of experiments with English L2 learners designed to determine to what extent the multiple distributional statistics of MWS inherent in register/genre-specific linguistic input (as estimated using a large corpus of actual language use) would affect the processing latencies of the MWS. We found the MWS frequency effect in three out of four registers investigated (all with the exception of fiction), i.e. our participants responded faster to higher frequency MWS, even after controlling for the effects of substring frequency. The finding that our participants showed MWS frequency effects in the spoken register is in line with the results of previous studies on adult native speakers and non-native speakers (e.g., Arnon & Snider, 2010; Tremblay et al., 2012; Hernandez et al. 2016). Importantly, our findings extend this

[5] Two anonymous reviewers recommended to pool the data from the four experiments and report on the interaction effects between our key predictors (log MWS frequency, slot entropy) and a 'register' variable. We have computed such a 'global' using orthogonal contrasts for the 'register' variable. This model revealed a significant effect of log MWS frequency ($= -0.026$, $SE = 0.006$, $t = -4.169$) but no significant interactions between log MWS frequency and register. Since we aimed to test whether our participants can detect and adapt to the changing statistics of multiple input types, we decided to report on four separate models in the study.

[6] To avoid overfitting resulting from multiple substring frequency control variables, we followed the procedure used in Arnon & Snider (2010) and first ran a model with all substring frequency controls and then removed the variables whose standard error was greater than the value of their coefficient in the model. All reported models had low collinearity (all $VIFs < 1.8$).

[7] Standardized beta coefficients indicate how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable. Pseudo-R2 for generalized mixed-effect models (GLMM) can be categorized into two types: Marginal R^2 represents the variance explained by fixed factors. Conditional R^2 is interpreted as variance explained by both fixed and random factors (i.e. the entire model).

Table 2: Results from the mixed effects regression models fitted to the data from the four experiments.

	Register comparison			
	Log. Reaction Time			
	Spoken	Fiction	Academic	News
Constant	0.286^{**}	0.752^{***}	0.447^{***}	0.532^{***}
	(0.069, 0.503)	(0.469, 1.036)	(0.302, 0.592)	(0.300, 0.763)
log MWS frequency	$B = -0.037^{**}$	$B = -0.013$	$B = -0.025^{**}$	$B = -0.035^{**}$
	(−0.061, −0.013)	(−0.039, 0.013)	(−0.042, −0.008)	(−0.060, −0.009)
	$\beta = -0.13$	$\beta = -0.06$	$\beta = -0.11$	$\beta = -0.13$
slot entropy (low to mid)	$B = 0.009$	$B = 0.019$	$B = -0.046^{**}$	$B = 0.037$
	(−0.066, 0.085)	(−0.039, 0.078)	(−0.078, −0.014)	(−0.012, 0.086)
	$\beta = 0.01$	$\beta = 0.03$	$\beta = -0.08$	$\beta = 0.11$
slot entropy (low to high)	$B = 0.069$	$B = -0.010$	$B = 0.006$	$B = -0.0003$
	(−0.013, 0.152)	(−0.067, 0.046)	(−0.029, 0.041)	(−0.050, 0.049)
	$\beta = 0.1$	$\beta = 0.02$	$\beta = -0.02$	$\beta = 0.04$
log MWS freq.:entropy (mid)	$B = -0.013$			
	(−0.044, 0.017)			
log MWS freq.:entropy (high)	$B = -0.041^{*}$			
	(−0.081, −0.001)			
log final bigram	$B = 0.002$	$B = 0.005$	$B = 0.004$	$B = -0.003$
	(−0.007, 0.011)	(−0.006, 0.016)	(−0.003, 0.011)	(−0.012, 0.005)
log final word	$B = 0.011$	$B = -0.031^{**}$	$B = -0.020^{***}$	$B = -0.010$
	(−0.002, 0.024)	(−0.050, −0.011)	(−0.031, −0.009)	(−0.027, 0.007)
length (char)	$B = 0.003$	$B = 0.002$	$B = 0.008^{***}$	$B = 0.008^{***}$
	(−0.004, 0.010)	(−0.005, 0.009)	(0.005, 0.012)	(0.003, 0.013)
pair variant (low to high)	$B = -0.065^{*}$	$B = 0.008$	$B = -0.028$	$B = -0.036$
	(−0.118, −0.012)	(−0.055, 0.071)	(−0.070, 0.014)	(−0.093, 0.022)
block order	$B = -0.025$	$B = -0.097^{***}$	$B = 0.010$	$B = -0.062^{***}$
	(−0.060, 0.009)	(−0.140, −0.054)	(−0.016, 0.035)	(−0.097, −0.028)
Conditional R^2	0.41	0.39	0.30	0.40
Marginal R^2	0.02	0.03	0.03	0.05

Note: *p<0.05; **p<0.01; ***p<0.001
Numbers in parentheses indicate 95% confidence intervals.
B indicate unstandardized beta coefficients
β indicate standardized beta coefficients

prior research by demonstrating that language users' ability to track the statistics in the input is not limited to the spoken conversational language but it can be observed in written registers/genres. In addition to the MWS frequency effects, the significant main effect of entropy found in the register of academic writing indicated that our participants were able to develop sensitivity to more complex distributional statistics, i.e. they showed faster response latencies with higher slot entropy. The direction of the entropy effect is consistent with the finding of Matthews & Bannard's (2010) study demonstrating that 2-3 years old children were more accurate in repeating MWS with higher slot entropy. Additional support for the facilitatory effect of more complex distributional statistics on the processing of MWS comes from the significant interaction between frequency and entropy indicating that the effect of MWS frequency increased with increasing degrees of MWS entropy.

To our knowledge, this is the first study to show that language users (whether native or non-native speakers) are able to tune to the multiple distributional statistics inherent in register-specific input types within a single language. The findings from this study provide a key contribution to a growing body of research that explore statistical learning through the lens of multilingual acquisition. This research has explored the consequences of accruing statistics in multilanguage input and has typically been conducted using artificial stimulus-sequences (cf., Bulgarelli, Lebkuecher & Weiss, 2018, for a recent overview). Our study has demonstrated how the acquisition of multiple statistics can be investigated on the basis of stimuli constructed from large corpora of au-

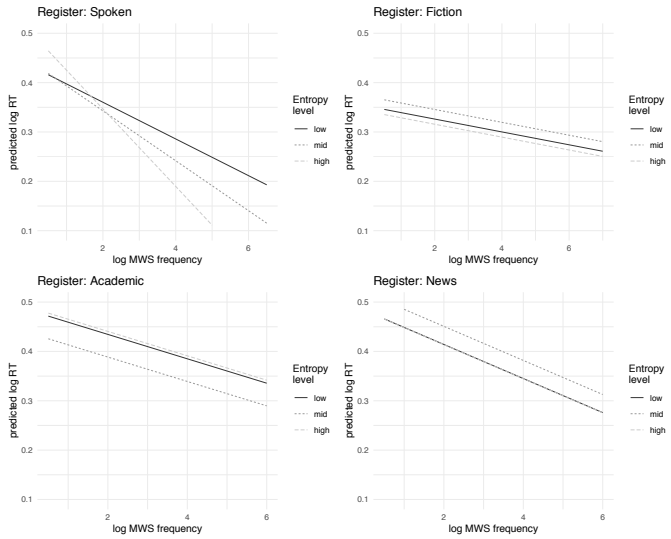

Figure 2: Effects of log MWS frequency by entropy level (high, mid, low) across registers.

thentic language data.

Some of the questions left open by the current study may provide interesting avenues for future work. First, we investigated sensitivity to the register-specific multiple statistics in adult second language learners. The question arises whether similar results could be obtained for adult native speakers. Second, in the light of the lifelong nature of language learning, it is of special importance important to track the developmental progression in response to the changes in the distributional properties of the linguistic input across the lifespan. This involves understanding not only the developmental progression during early stages of child language acquisition (prior to literacy) but also understanding the nature of such progression during later stages of language development, which is strongly driven by the distributional statistics of written input (Seidenberg & MacDonald, 2018). And, third, it would be important to determine whether the ability to tune to multiple statistics is subject to individual variability, and if so, to what extent this variability is linked to other cognitive, affective and environmental factors.

References

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining l1–l2 differences. *Topics in Cognitive Science*, *9*(3), 621–636.

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

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*(3), 241–248.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Grothendieck, G. (2015). Package 'lme4'. *Convergence*, *12*(1).

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(s1), 1–26.

Biber, D. (1999). A register perspective on grammar and discourse: variability in the form and use of english complement clauses. *Discourse Studies*, *1*(2), 131–150.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, *26*, 181–190.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Bulgarelli, F., Lebkuecher, A. L., & Weiss, D. J. (2018). Statistical learning and bilingualism. *Language, speech, and hearing services in schools*, *49*(3S), 740–753.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234.

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, *6*(5), 259–278.

Chater, N., & Christiansen, M. H. (2018). Language acquisition as skill learning. *Current Opinion in Behavioral Sciences*, *21*, 205–208.

Christiansen, M. H., & Chater, N. (2016a). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2016b). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39, e62*.

Daelemans, W., & Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge University Press.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17–44.

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785–800.

Kerz, E., & Wiechmann, D. (2017). Individual differences in L2 processing of multi-word phrases: Effects of working memory and personality. In R. Mitkov (Ed.), *Computational and corpus-based phraseology. EUROPHRAS 2017* (pp. 306–321).

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, 154–169.

Lemhöfer, K., & Broersma, M. (2012). Introducing lextale: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, *44*(2), 325–343.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the Association for Computational Linguistics: system demonstrations* (pp. 55–60).

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. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8-and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447-468.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181-203.

Seidenberg, M. S., & MacDonald, M. C. (2018). The impact of language experience on language and reading: A statistical learning approach. *Topics in Language Disorders*, *38*(1), 66–83.

Shaoul, C., & Westbury, C. (2011). Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon*, *6*(1), 171–196.

Team, R. C., et al. (2013). R: A language and environment for statistical computing.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516.

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, *33*(4), 519–549.