Cognition 148 (2016) 79–84

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Short Communication

Effects of statistical learning on the acquisition of grammatical categories through Qur'anic memorization: A natural experiment

Fathima Manaar Zuhurudeen, Yi Ting Huang *

University of Maryland College Park, United States

ARTICLE INFO

Article history:
Received 22 September 2014
Revised 17 December 2015
Accepted 23 December 2015
Available online 30 December 2015

Keywords:
Statistical learning
Grammatical categories
Language acquisition
Arabic

ABSTRACT

Empirical evidence for statistical learning comes from artificial language tasks, but it is unclear how these effects scale up outside of the lab. The current study turns to a real-world test case of statistical learning where native English speakers encounter the syntactic regularities of Arabic through memorization of the Qur'an. This unique input provides extended exposure to the complexity of a natural language, with minimal semantic cues. Memorizers were asked to distinguish unfamiliar nouns and verbs based on their co-occurrence with familiar pronouns in an Arabic language sample. Their performance was compared to that of classroom learners who had explicit knowledge of pronoun meanings and grammatical functions. Grammatical judgments were more accurate in memorizers compared to non-memorizers. No effects of classroom experience were found. These results demonstrate that real-world exposure to the statistical properties of a natural language facilitates the acquisition of grammatical categories.

1. Introduction

Linguistic representations form co-occurrence patterns that can be readily exploited during acquisition. Statistical learning is argued to play a central role in word segmentation and learning (Saffran, Aslin, & Newport, 1996; Pelucchi, Hay, & Saffran, 2009a, 2009b; Hay, Pelucchi, Graf Estes, & Saffran, 2011; Lew-Williams, Pelucchi, & Saffran, 2011; Thiessen & Saffran, 2003). It is also considered essential for acquiring grammatical rules, which operate over abstract categories that are not explicitly stated in the input (Cartwright & Brent, 1997; Harris, 1951; Maratsos & Chalkley, 1980). For example, a child who hears phrases like "*her cat*," "*her bike*," and "*her train*" can use distribution cues to infer that words which follow possessive pronouns form a category of nouns. Empirical evidence for this process comes primarily from studies that simulate real-world acquisition through artificial language tasks in children and adults (Gomez & Gerken, 2000; Hudson Kam & Newport, 2005; Langus, Marchetto, Bion, & Nespor, 2012; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Mintz, 2006; Misyak & Christiansen, 2012; Thompson & Newport, 2007; Wonnacott, Newport, & Tanenhaus, 2008).

Nevertheless, there remain important questions about how experimental findings extend to development outside the lab. Natural languages offer diverse cues to learning that may converge or conflict with distributional statistics. Consequently, it is unclear whether learners still detect statistical cues in the face of increased input complexity or prefer to learn from other cues. Recent research has found that when transitional probability and input quantity were sufficiently high (Hay et al., 2011; Lew-Williams et al., 2011), English-learning infants rely on statistical cues to segmentation words in an Italian language sample (Pelucchi et al., 2009a, 2009b). However, studies pitting statistical vs. prosodic cues have found preferences for the latter among infants (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Thiessen & Saffran, 2003) and adults (Langus et al., 2012). Infants are also less sensitive to statistical cues when utterances vary in length (Johnson & Tyler, 2010).

Real-world language acquisition also differs in the sheer quantity and duration of learning. Given limitations of participant attention and experimenter resources, artificial language tasks often involve short input-exposure durations (e.g., less than 20 min) and assess learning immediately after familiarization. Even studies that measure later retention typically do so within hours or days of initial exposure (Apfelbaum, Hazeltine, & McMurray, 2012; Arciuli & Simpson, 2012; Hudson Kam & Newport, 2005; Kim, Seitz, Feenstra, & Shams, 2009; Thompson & Newport, 2007; Wonnacott et al., 2008). To examine long-term impacts of statistical learning, recent studies have taken an individual-differences approach. Performance in statistical learning tasks has been shown to predict language outcomes in adults (Conway, Baurnschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2012), children

* Corresponding author at: Department of Hearing and Speech Sciences, University of Maryland College Park, 0100 Lefrak Hall, College Park, MD 20742, United States.

E-mail address: ythuang1@umd.edu (Y.T. Huang).

http://dx.doi.org/10.1016/j.cognition.2015.12.014
0010-0277/

(Arciuli & Simpson, 2012; Kidd, 2012), and impaired populations (Evans, Saffran, & Robe-Torres, 2009; Tomblin, Mainela-Arnold, & Zhang, 2007). Nevertheless, it can be difficult to isolate causal relationships since key measures involve substantial differences in content and task demands (e.g., visual-sequences learning, standardized grammar tests).

The current study takes a different approach by investigating a real-world test case of statistical learning. Many Muslims living in the US acquire Arabic as a native language at home or as a second language in the classroom, but there is a significant number who receive neither input exposure. These individuals are native English-speakers. However, since their families emigrate from countries outside of the Middle East (e.g., Sri Lanka, India, Malaysia, Somalia), they have limited access to an Arabic-speaking community. Nevertheless, they encounter the statistical regularities of Arabic through their memorization of the Qur'an, the primary religious text of Islam. This practice starts as early as four years of age, occurs for several hours a day, and continues for many years. Much like artificial language tasks, input of this kind rarely provides direct translations or topic discussion. Thus, these contexts isolate sensitivity to statistical cues without semantic confounds. They also offer unique opportunities to assess long-term impacts of statistical learning within a natural language.

The current study familiarized memorizers to a brief Arabic language sample featuring frequent closed-class words (subject/possessive pronouns) and infrequent open-class words (nouns/verbs). Memorizers then made grammatical judgments requiring categorization of open-class words based on co-occurrence with closed-class words. We compared their performance to classroom learners who had explicit knowledge of word meanings and syntactic functions. If experience with Qur'anic memorization generated knowledge of the transitional probabilities of closed-class words, then grammatical judgments of the current language sample may be more accurate in memorizers than classroom learners. If, however, learning based on prior statistical input was limited due to the complexity of natural input or lack of semantic cues, then accuracy in memorizers may be lower than classroom learners.

2. Method

2.1. Participants

Fifty-two participants took part in this study. From this group, data were excluded because of participant ($n = 2$) or experimenter ($n = 2$) errors. This resulted in a sample of 48 individuals who were recruited based on a 2×2 design. *Memorization* compared individuals who engaged in Qur'anic memorization to those who did not. *Classroom* compared individuals who took Arabic language class to those who did not. Both memorizers and classroom learners were required to have at least one semester's worth of Arabic experience and remain active at the time of testing. Memorizers were recruited from mosques and Islamic centers, in the Washington, D.C. metro area. Approximately 95% of respondents listed English as their primary mode of communication. Importantly, in cases where this was not true, they always indicated a non-Arabic language (e.g., Urdu, Bengali, Somali). Non-memorizers were recruited from the Muslim Students Association and the Arabic Studies Department at the University of Maryland. All participants identified themselves as non-native Arabic speakers.

To confirm differences in Arabic experience across groups, participants completed the Bilingual Language Profile: English–Arabic after the primary task (Birdsong, Gertken, & Amengual, 2012). This survey assessed quantity of prior exposure (in years), current weekly exposure (in hours), and self-rated proficiency (on 0–6 scale). Participants also translated the current stimuli and identified their parts of speech (out of 12 items). These data were analyzed through a series of linear models, using the lme4 software package in R (Bates, 2007). Analyses confirmed effects of memorization and classroom learning in the current sample (Table 1). Relative to non-memorizers, memorizers had more prior exposure ($t = 10.01$, $p < .001$) and higher self-rated proficiency ($t = 10.01$, $p < .001$). Similarly, relative to non-classroom learners, classroom learners had more prior exposure ($t = 2.56$, $p < .05$), current exposure ($t = 2.61$, $p < .05$), and higher self-rated proficiency ($t = 10.30$, $p < .001$). Classroom effects on ratings were greater in non-memorizers compared to memorizers, leading to an additional interaction ($t = 2.08$, $p < .05$). Critically, measures also revealed key differences among classroom learners and memorizers. Relative to non-classroom learners, classroom learners translated ($t = 5.00$, $p < .001$) and identified parts of speech for more items ($t = 3.51$, $p < .01$). However, memorizers did not differ from non-memorizers in their translations (p's > .30) and were *less* accurate at identifying parts of speech ($t = 2.45$, $p < .05$). This confirmed that unlike classroom learners, memorizers had limited explicit knowledge of Arabic.

2.2. Materials and procedures

During the familiarization phase, participants were told to listen to a 5-min sample of Arabic sentences. Sentences consisted of open-class categories (nouns/verbs) and closed-class categories (subject/possessive pronouns). Items from closed-class categories were monosyllabic and highly frequent while open-class items were bisyllabic and highly infrequent (Table 2). Analyses confirmed that open-class items were often unfamiliar to participants (Table 1). Items were combined to create eight unique sentences based on Arabic syntax: (1) subject pronouns occurring after verbs (e.g., A_2B_1: "*farar-tu*" means I FLED) and (2) possessive pronouns after nouns (e.g., C_1D_1: "*dalwa-ha*" means HER BUCKET). Sentences were repeated 23 times in a semi-randomized order. To allow for tests of generalization, each open-class item was paired with only one closed-class item within an order list (e.g., A_1B_1 but not A_1B_2).

During the test phase, each trial featured a pair of grammatical and ungrammatical phrases. Participants first heard phrases presented sequentially, with the order of presentation randomized across trials. They were then asked to select the phrase that sounded grammatical and to guess if necessary. Across trials, grammatical phrases featured two-word combinations that appeared during the familiarization phase (*Familiarity Test*) or novel combinations from the same stock of words (*Generalization Test*). These phrases were paired with ungrammatical phrases that either: (1) repeated tokens from the same category, e.g., C_1C_2 (*Repetition trials*); (2) reversed positions of within-phrase categories, e.g., B_1A_1 (*Reversal trials*); or (3) replaced pairings of open- and closed-class categories, e.g., A_1D_1 (*Replacement trials*). Thus, generalization in Replacement trials provided a critical test of whether categories of open-class items were formed since judgments could not be based on familiarity or explicit knowledge. Eight tokens of each type were randomly presented in the Familiarity and Generalization Tests.

All stimuli were pre-recorded by a female, native Arabic speaker. To limit acoustic cues to phrase boundaries, familiarization sentences were carefully spoken with a consistent tempo and limited prosody (list intonation). Subsequent analysis revealed no significant differences in pitch contour, stress, vowel duration, and pauses between words that occurred within and between phrases (all p's > .05). Analyses of test phase stimuli also confirmed no significant differences between grammatical and ungrammatical phrases across trial types (all p's > .05). Two order lists counterbalanced the category combinations presented in familiarization and test phases. See Appendices A and B for a full list of familiarization and test items.

Table 1
Language experience by participant group.

	Prior exposure (in years)[a,b]	Current exposure (hrs/week)[b]	Meaning translation (out of 12)[b]	Parts of speech ID (out of 12)[a,b]	Self-rated proficiency (0–6)[a,b,c]
Memorizers with classroom	***M* = 9.0** SD = 3.7 Range: 4–17	***M* = 9.3** SD = 17.8 Range: 1–65	***M* = 2.2** SD = 3.2 Range: 0–10	***M* = 3.4** SD = 3.1 Range: 0–11	***M* = 3.5** SD = 0.8 Range: 2.2–4.5
Memorizers without classroom	***M* = 7.8** SD = 3.1 Range: 2–13	***M* = 5.3** SD = 8.1 Range: 0–24	***M* = 0.1** SD = 0.3 Range: 0–1	***M* = 1.5** SD = 2.4 Range: 0–7	***M* = 1.5** SD = 1.1 Range: 0–3.2
Non-memorizers with classroom	***M* = 2.5** SD = 1.0 Range: 1–5	***M* = 21.6** SD = 27.7 Range: 3–84	***M* = 3.2** SD = 1.7 Range: 0–5	***M* = 6.7** SD = 2.3 Range: 4–10	***M* = 3.0** SD = 0.9 Range: 1.5–4
Naïve listeners	***M* = 0** SD = 0 Range: 0	***M* = 0** SD = 0 Range: 0	***M* = 0.1** SD = 0.3 Range: 0–1	***M* = 2.5** SD = 3.9 Range: 0–11	***M* = 0** SD = 0 Range: 0

[a] Indicates main effect of memorization.
[b] Indicates main effect of classroom.
[c] Indicates interaction between memorization and classroom.

Table 2
Translation and frequency of open- (A and C) and closed-class (B and D) items.

Word token	Grammatical category			
	A (*verb*)	**B** (*subj. pronoun*)	**C** (*noun*)	**D** (*poss. pronoun*)
1	**baTash** English: SEIZED Freq: 0.2	**tu** English: I Freq: 504.1	**dalwa** English: BUCKET Freq: 0.1	**ha** English: HER/ITS Freq: 1799.4
2	**farar** English: FLED Freq: 0.2	**tum** English: YOU (pl.) Freq: 888.7	**baqla** English: HERBS Freq: 0.1	**ki** English: YOUR (f.) Freq: 888.7
3	**awey** English: RETIRED Freq: 0.1	–	**jidh'a** English: TREE TRUNK Freq: 0.2	–
4	**kanaz** English: HOARDED Freq: 0.1	–	**ma'waa** English: ABODE Freq: 2.8	–

Note: Frequency estimates were obtained through the Qur'an Arabic corpus and are calculated as number of occurrences out of 10,000 words (Dukes, 2011; Sharaf, 2011).

3. Results

3.1. Comparisons of group performance

We examined the accuracy of grammaticality judgments using a series of logistic mixed-effects models (Jaeger, 2008). Memorization and classroom were modeled as fixed effects while subjects/items were modeled as random effects on the intercept only. Fig. 1 illustrates high overall accuracy in the Familiarity Test (M = 88%, SD = 12%). In Repetition and Reversal trials, no effects of memorization or classroom were found (all p's > .30). However, in Replacement trials, accuracy was higher in memorizers compared to non-memorizers (84% vs. 73%; z = 2.21, p < .05). This suggests that prior experience with the statistical properties of Arabic increased sensitivity to the co-occurrence of open- and closed-class items in the current study. The fact that this advantage emerged specifically in Replacement trials suggests that memorization did not lead to general improvements in grammatical judgements. Instead, advantages were isolated to situations where accuracy depended on statistically-motivated categorizations. No effects of classroom were found (all p's > .90).

However, it was possible that accurate judgements by memorizers were based on recall of previously-encountered phrases rather than categorizations of the items. Since memorizers have extensive experience memorizing Arabic utterances, they could have rejected unfamiliar/ungrammatical phrases in the test phase on this basis. If this were true, then advantages for memorizers should be absent in the Generalization Test. If, however, memorizers used statistical cues to categorize open-class items, then they should continue to exhibit higher accuracy than non-memorizers, even with novel grammatical phrases.

Fig. 2 illustrates high overall accuracy in the Generalization Test (M = 85%, SD = 11%). No effects of memorization or classroom were found in Repetition trials (all p's > .20). There was a marginal benefit of classroom experience in Reversal trials (96% vs. 89%; z = 1.81, p < .10). This may reflect the fact that classroom learners often explicitly know the relative position of closed- and open-class items, which are distinguished by syllable length in the current task. Critically, when this strategy was unavailable in Replacement trials, no effects of classroom were found (all p's > .30). Nevertheless, similar to the Familiarity Test, accuracy was higher in memorizers compared to non-memorizers (81% vs. 69%; z = 2.28, p < .05). This demonstrates that memorizers possessed implicit knowledge of the statistical regularities of Arabic pronouns and used this to generate accurate inferences about the categories of unfamiliar nouns/verbs.

3.2. Predictors of individual performance

Finally, memorizers in the current study had more years of Arabic experience compared to classroom learners (8.4 vs. 5.7 years on average). This raises the possibility that effects of the former but not the latter reflected overall input quantity rather than specific experience type. To distinguish the two, we examined relationships between measures of language experience (Table 3). Self-rated proficiency was correlated with prior and current exposure, meaning translation, and part of speech identification (all r's > .30, all p's < .05). Similarly, current exposure was correlated with translation and speech identification (all r's > .30, all p's < .05). However, the same was not true of prior exposure (all r's < .20, all p's > .30). This suggests that prior experience did not always generate more explicit linguistic knowledge. Next, we examined relationships between language experience and average accuracy in the Familiarity and Generalization Tests. Accuracy in Repetition and Reversal trials was unrelated to all measures (all r's < .20, all p's > .15). Similarly, accuracy in Replacement trials was unrelated to self-rated proficiency, translation, and speech identification (all r's < .15, all p's > .50). This demonstrates that

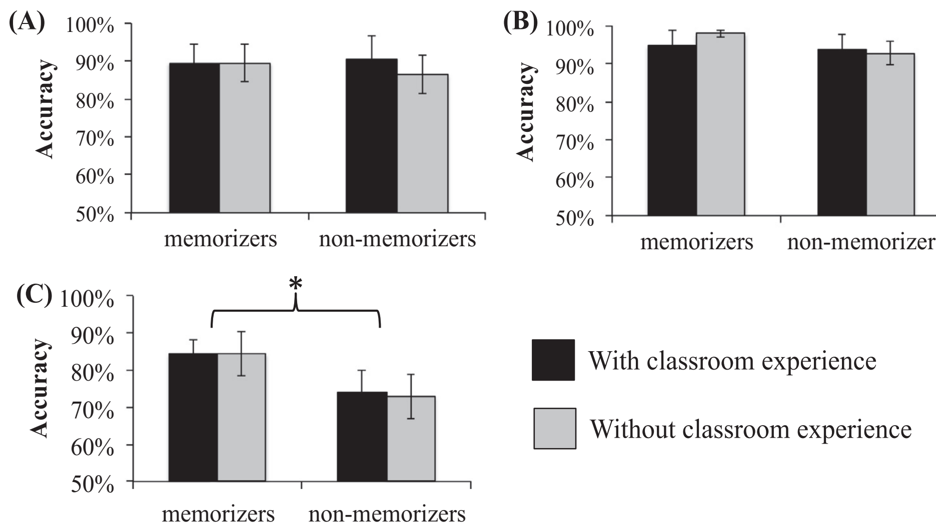

Fig. 1. In the Familiarity Test, effects of memorization and classroom on accuracy in the (A) Repetition trials, (B) Reversal trials, and (C) Replacement trials. Note: $*$ denotes significant differences at $p < .05$ level.

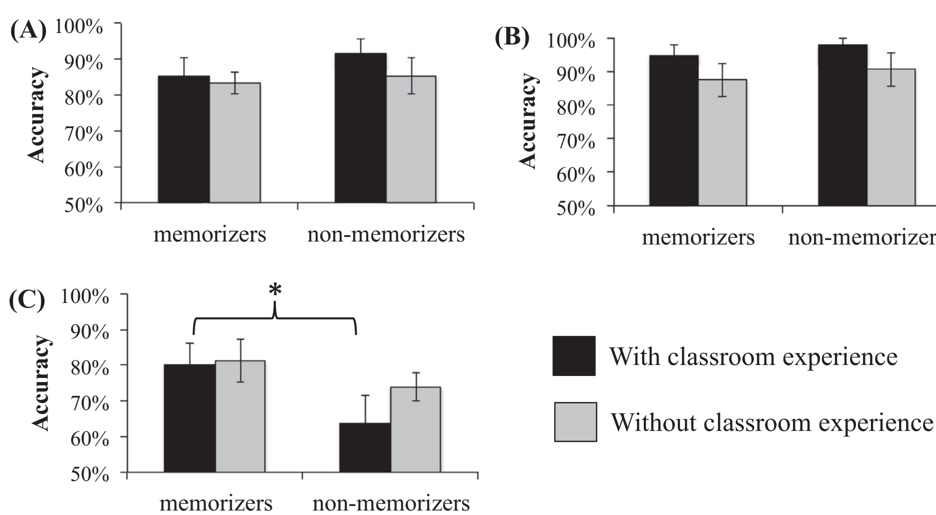

Fig. 2. In the Generalization Test, effects of memorization and classroom on accuracy in the (A) Repetition trials, (B) Reversal trials, and (C) Replacement trials. Note: $*$ denotes significant differences at $p < .05$ level.

Table 3
In Replacement trials, correlations between language experience and accuracy.

	Prior exposure	Current exposure	Meaning translation	Part of speech ID	Self-rated proficiency
Current exposure	−0.07				
Meaning translation	0.06	0.59**			
Part of speech ID	−0.14	0.37*	0.64**		
Self-rated proficiency	0.43**	0.35*	0.46**	0.32*	
Accuracy	0.34*	−0.29*	0.10	0.01	0.09

* Denotes significant at $p < .05$ level.
** Denotes significant at $p < .01$ level.

measures of explicit language experience did not always predict accuracy in the current task.

However, accuracy in Replacement trials was correlated with prior and current Arabic exposure. On their own, these effects are difficult to interpret since memorizers and classroom learners differed both in their type *and* amount of experience. To distinguish these effects, we conducted separate correlations of input quantity within participants of similar experience type. Prior exposure varied within memorizer, but this was unrelated to accuracy ($r(22) = .23$, $p > .20$). In contrast, current exposure varied within non-memorizers and was *negatively* correlated with accuracy ($r(22) = -.44$, $p < .05$). Since non-memorizers acquired experience through classroom instruction, this suggests that explicit syntactic knowledge may interfere with sensitivity to statistical cues. Remaining measures of language experience were unrelated to accuracy (all r's $< .20$, all p's $> .20$). These results suggest that advantages in grammatical judgments by memorizers reflect effects of experience type and not overall input quantity.

4. Discussion

The current study examined grammatical knowledge in two populations with distinct language experiences: (1) Qur'an memorizers, who regularly encounter the statistical properties of Arabic and (2) classroom learners, who have explicit knowledge of words and rules. We found that while all participants were sensitive to co-occurrence patterns within a brief Arabic language sample, inferences of grammatical categories were more accurate in memorizers compared to non-memorizers. In contrast, effects of classroom experience were minimal. These results suggest that prolonged exposure to the statistical properties of a natural language can facilitate acquisition of grammatical categories. This work provides an important bridge between empirical studies of statistical learning in the lab and language acquisition in the wild.

One potential concern was that effects of memorization may have reflected sensitivity to the prosodic properties of Arabic, rather than statistical learning. For example, memorizers may have categorized words via cues to phrasal boundaries or rejected ungrammatical phrases based on infelicitous prosody. Prosodic bootstrapping features prominently in accounts of syntactic development (Gleitman & Wanner, 1982; Morgan & Newport, 1981). Nevertheless, it does not explain judgments in the current study since familiarization sentences provided no reliable prosodic cues. Even when grammatical and ungrammatical phrases were distinguished via syllable length in Reversal trials, there was no evidence that memorizers recruited this cue. Instead, their advantage occurred specifically in critical Replacement trials, when phrases were matched for prosody and judgments required categorizations based on statistical cues.

One surprising feature of our findings was the limited benefit of classroom experience in the current task. While classroom learners often provided accurate translations of pronouns and identification of their grammatical categories, this knowledge did not increase sensitivity to co-occurrence patterns in the language sample. In fact, we found that more weekly exposure to Arabic led to less accurate judgments in Replacement trials, suggesting that explicit knowledge of the grammatical functions of closed-class items may have interfered with statistical learning. These results have implications for educational curricula in second language acquisition. Teaching methods traditionally rely on metalinguistic knowledge of the first language to acquire grammatical features in the second (e.g., introducing "*tum*" as the plural form of the second person pronoun). Yet, our findings suggest that this top-down approach may negatively impact learners' sensitivity to the bottom-up statistics of a language.

Our results also raise important questions about the nature of the input for memorizers and the precise statistical cues that are used to form syntactic representations. These variables are easily manipulated within artificial language tasks. However, they are less transparent in the current study since memorization involves interactions with a natural language over several years. Moreover, it remains unknown how much input experience is necessary to generate reliable categories. Among our current sample, we found that grammatical judgments were unaffected by input quantity, as measured by years of prior exposure and current weekly exposure. However, since all our participants had at least two years of memorization experience and averaged seven hours of weekly exposure, it may be the case that quantity effects would be observed in less experienced populations or in more challenging tasks.

Finally, there remain many questions about the nature of the linguistic knowledge acquired by memorizers. Memorizers rarely provided translations of the current stimuli, confirming that their input offers minimal semantic information. Nevertheless, their ability to distinguish nouns and verbs suggests that they had formed distinct categories of pronouns. Future studies will examine the specificity and generality of syntactic representations in this population. Beyond the area of syntax, the types of phonological categories and lexical entries that are formed on the basis of statistical experiences remain unknown. These are enduring questions within the field of language acquisition. The current test case offers a unique situation where the input (Qur'an), context (memorization), and quantity (years) are well defined. This presents exciting opportunities for studying statistical learning in environments that both mimic the properties of artificial language tasks and approximate the complexities of real-world language acquisition.

Acknowledgments

We would like to thank Nan Bernstein Ratner, Yasmeen Faroqi Shah, and Alexa Romberg for their comments on an earlier draft of this paper. This work was supported by grants from the Mary Cobey Martin Fund for Research Excellence to FMZ and the ADVANCE Program for Inclusive Excellence to YTH.

Appendix A. List of familiarization items for Orders 1 and 2

Order	Arabic sentence	Category	Translation
1	baTashtu ma'waaki	$A_1\ B_1\ C_4\ D_2$	I seized your (f.) abode
	baTashtu baqlaki	$A_1\ B_1\ C_2\ D_2$	I seized your (f.) herbs
	farartum baqlaki	$A_2\ B_2\ C_2\ D_2$	You (pl.) fled your (f.) herbs
	farartum jidh'aha	$A_2\ B_2\ C_3\ D_1$	You (pl.) fled her tree trunk
	Aweytu jidh'aha	$A_3\ B_1\ C_3\ D_1$	I retired her tree trunk
	Aweytu dalwaha	$A_3\ B_1\ C_1\ D_1$	I retired her bucket
	Kanaztum dalwaha	$A_4\ B_2\ C_1\ D_1$	You (pl.) hoarded her bucket
	Kanaztum ma'waaki	$A_4\ B_2\ C_4\ D_2$	You (pl.) hoarded your (f.) abode
2	baTashtum ma'waaha	$A_1\ B_2\ C_4\ D_1$	You (pl.) seized her abode
	baTashtum dalwaki	$A_1\ B_2\ C_1\ D_2$	You (pl.) seized your (f.) bucket
	Farartu ma'waaha	$A_2\ B_1\ C_4\ D_1$	I fled her abode
	Farartu baqlaha	$A_2\ B_1\ C_2\ D_1$	I fled her herbs
	Aweytum baqlaha	$A_3\ B_2\ C_2\ D_1$	You (pl.) retired her herbs
	Aweytum jidh'aki	$A_3\ B_2\ C_3\ D_2$	You (pl.) retired your (f.) tree trunk
	Kanaztu jidh'aki	$A_4\ B_1\ C_3\ D_2$	I hoarded your (f.) tree trunk
	Kanaztu dalwaki	$A_4\ B_1\ C_1\ D_2$	I hoarded your (f.) bucket

Appendix B. List of test items for Orders 1 and 2

Test	Trial	Order 1 – Gram.	Order 1 – Ungram.	Order 2 – Gram.	Order 2 – Ungram.
Familiarity	Repetition	A_1 D_1	D_1 D_2	A_1 D_2	D_1 D_2
		B_1 C_1	B_1 B_2	B_1 C_2	B_1 B_2
		A_2 D_2	A_2 A_1	A_2 D_1	A_2 A_1
		B_3 C_2	C_1 C_2	B_3 C_1	C_1 C_2
		A_3 D_1	A_3 A_4	A_3 D_2	A_3 A_4
		B_4 C_2	B_3 B_4	B_4 C_1	B_3 B_4
		A_4 D_2	A_4 A_1	A_4 D_1	A_4 A_1
		B_2 C_1	B_2 B_4	B_2 C_2	B_2 B_4
	Reversal	A_1 D_1	D_1 A_1	A_1 D_2	D_2 A_1
		B_1 C_1	C_1 B_1	B_1 C_2	C_2 B_1
		A_2 D_2	D_2 A_2	A_2 D_1	D_1 A_2
		B_3 C_2	C_2 B_3	B_3 C_1	C_1 B_3
		A_3 D_1	D_1 A_3	A_3 D_2	D_2 A_3
		B_4 C_2	C_2 B_4	B_4 C_1	C_1 B_4
		A_4 D_2	D_2 A_4	A_4 D_1	D_1 A_4
		B_2 C_1	C_1 B_2	B_2 C_2	C_2 B_2
	Replacement	A_1 D_1	B_1 D_1	A_1 D_2	B_1 D_2
		B_1 C_1	A_1 C_1	B_1 C_2	A_1 C_2
		A_2 D_2	B_2 D_2	A_2 D_1	B_2 D_1
		B_3 C_2	A_3 C_2	B_3 C_1	A_3 C_1
		A_3 D_1	B_3 D_1	A_3 D_2	B_3 D_2
		B_4 C_2	A_4 C_2	B_4 C_1	A_4 C_1
		A_4 D_2	B_4 D_2	A_4 D_1	B_4 D_1
		B_2 C_1	A_2 C_1	B_2 C_2	A_2 C_2
Generalization	Repetition	A_1 D_2	D_1 D_2	A_1 D_1	D_1 D_2
		B_2 C_2	B_1 B_2	B_2 C_1	B_1 B_2
		A_2 D_1	A_2 A_1	A_2 D_2	A_2 A_1
		B_3 C_1	C_1 C_2	B_3 C_2	C_1 C_2
		A_3 D_2	A_3 A_4	A_3 D_1	A_3 A_4
		B_4 C_1	B_3 B_4	B_4 C_2	B_3 B_4
		A_4 D_1	A_4 A_1	A_4 D_2	A_4 A_1
		B_1 C_2	B_2 B_4	B_1 C_1	B_2 B_4
	Reversal	A_1 D_2	D_2 A_1	A_1 D_1	D_1 A_1
		B_2 C_2	C_2 B_2	B_2 C_1	C_1 B_2
		A_2 D_1	D_1 A_2	A_2 D_2	D_2 A_2
		B_3 C_1	C_1 B_3	B_3 C_2	C_2 B_3
		A_3 D_2	D_2 A_3	A_3 D_1	D_1 A_3
		B_4 C_1	C_1 B_4	B_4 C_2	C_2 B_4
		A_4 D_1	D_1 A_4	A_4 D_2	D_2 A_4
		B_1 C_2	C_2 B_1	B_1 C_1	C_1 B_1
	Replacement	A_1 D_2	B_1 D_2	A_1 D_1	B_1 D_1
		B_2 C_2	A_2 C_2	B_2 C_1	A_2 C_1
		A_2 D_1	B_2 D_1	A_2 D_2	B_2 D_2
		B_3 C_1	A_3 C_1	B_3 C_2	A_3 C_2
		A_3 D_2	B_3 D_2	A_3 D_1	B_3 D_1
		B_4 C_1	A_4 C_1	B_4 C_2	A_4 C_2
		A_4 D_1	B_4 D_1	A_4 D_2	B_4 D_2
		B_1 C_2	A_1 C_2	B_1 C_1	A_1 C_1

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2015.12.014.

References

Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2012). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology, 49*, 1348–1365.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience, 517*, 133–135.

Bates, D. M. (2007). *Linear mixed model implementation in lme4*. Madison: Manuscript, University of Wisconsin (January 2007).

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. COERLL University of Texas at Austin. Retrieved from <https://sites.la.utexas.edu/bilingual/>.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*, 121–170.

Conway, C. M., Baurnschmidt, A., Huang, S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356–371.

Dukes, K. (2011). *The Qur'an Arabic corpus*. Retrieved from <http://corpus.quran.com/>.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321–335.

Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, UK: Cambridge University Press.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago, IL: University of Chicago Press.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*, 93–106.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

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.

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

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*, 131–141.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339–345.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171–184.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*, 145–149.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66*, 285–306.

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*, 1323–1329.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.). *Children's language* (Vol. 2, pp. 127–214). New York: Gardner Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Mintz, T. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets words: How children learn verbs* (pp. 31–63). Oxford, England: Oxford University Press.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20*, 67–85.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*, 244–247.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*, 674–685.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Sharaf, A. (2011). *Text mining the Qur'an*. Retrieved from <http://www.textminingthequran.com/>.

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

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*, 1–42.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without Specific Language Impairment. *Language Learning and Development, 3*, 269–293.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165–209.