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.