Distributional Effects and Individual Differences in L2 Morphology Learning

Patricia J. Brooks, Nicole Kwoka, Vera Kempe

Distributional Effects and Individual Differences in L2 Morphology Learning

Patricia J. Brooks

College of Staten Island and The Graduate Center, CUNY

Nicole Kwoka

College of Staten Island, CUNY

and

Vera Kempe

Abertay University Dundee

Key words:

miniature natural language learning, inflectional morphology, statistical learning, nonverbal

intelligence, entropy, second language learning

Address correspondence to:

Patricia Brooks

Department of Psychology, College of Staten Island, CUNY

2800 Victory Blvd., 4S-108, Staten Island, NY 10314

patricia.brooks@csi.cuny.edu

Abstract

Foreign (L2) language-learning outcomes may depend on the structure of the input as well as learners' cognitive abilities. This study tested whether less predictable input might facilitate learning and generalization of L2 morphology, while evaluating contributions of statistical-learning ability, nonverbal intelligence, phonological short-term memory and verbal working-memory. Over three sessions, adults (N=54) were exposed to a Russian case-marking paradigm with either a balanced or skewed distribution of items in the input. Whereas both statistical-learning ability and nonverbal intelligence predicted learning of trained items, only nonverbal intelligence also predicted generalization of case-marking inflections to new vocabulary. Neither measure of temporary-storage capacity predicted learning. Balanced, less predictable input was associated with higher accuracy in generalization, but only in the initial test session. These results suggest that individual differences in pattern-extraction abilities play a more sustained role in L2 acquisition than instructional manipulations that vary the predictability of lexical items in the input.

Distributional Effects and Individual Differences in L2 Morphology Learning

Many richly inflected languages, such as Russian, Lithuanian, or Hungarian, contain complex case-marking systems wherein nouns are categorized into declensions based on grammatical features, such as gender or number. Adult L2 learners often find it challenging to form grammatical categories on the basis of distributional information, such as when using different sets of case markers, articles, and pronouns to distinguish masculine from feminine nouns (DeKeyser, 2005). In the interest of developing effective L2 pedagogy, it is important to examine whether the input might be structured in ways that help learners to notice distributional cues to grammatical categories, while bearing in mind other known variables that impact L2 learning trajectories. In practice, establishing the efficacy of an input manipulation may be difficult in light of considerable individual differences in language-learning aptitude, which to a large extent are attributable to variation in basic cognitive abilities, such as phonological short-term memory (Baddeley, Gathercole, & Papagno, 1998; Gupta & MacWhinney, 1997; Service, 1992), verbal working memory capacity (Linck, Osthus, Koeth, & Bunting, 2014; Miyake & Friedman, 1998; Service, 1992), statistical-learning ability (Granena, 2013, Linck et al., 2013; Speciale, Ellis, & Bywater, 2004) and nonverbal intelligence (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012; Grigorenko, Sternberg, & Ehrman, 2000). The current study aimed to evaluate the effect of varying the distribution of nouns instantiating a case-marking paradigm on L2 learning of inflectional morphology while bearing in mind individual differences in language-learning aptitude.

Input manipulations that benefit L2 morphology learning

Acquiring a case-marking paradigm involves learning to associate sets of inflectional morphemes with specific semantic and syntactic contexts. For example, in the Russian phrase *k*

stolu [towards table], the dative suffix *–u* indicates that the noun *stol* [table] takes on the role of the goal of motion in the context of the preposition *k* [towards]. Learners may register how specific nouns are inflected based on their occurrence in the input, e.g., by hearing forms like *stolu* [table-DAT] or *divanu* [couch-DAT] in the context of the preposition *k* [towards]. They may notice that any given noun is associated with a set of inflections—e.g., that *stol* takes *–u* in dative, *–a* in genitive, and *–om* in instrumental case—and may come to recognize that sets of inflections reflect grammatical category membership—i.e., all nouns that take *–u* in dative, *–a* in genitive, and *–om* in instrumental case belong to the category of masculine nouns. Having identified underlying grammatical categories, learners may apply inflectional morphemes productively, e.g., identifying *divan* as masculine enables learners to combine it correctly with the instrumental suffix *–om* even if the inflected form *divanom* has never been encountered.

Two types of input manipulation have been shown to facilitate L2 morphology learning. First, providing redundant form-based cues that distinguish members of gender categories appears to facilitate learning (Gerken, Wilson, & Lewis, 2005). In a series of experiments, American infants of ages 16 to 17 months were exposed to a miniature Russian case-marking paradigm consisting of masculine and feminine nouns inflected for accusative (ACC) and instrumental (INS) case. In each gender category, a subset of the nouns shared a form-based feature, i.e., *tel'* for masculine nouns (e.g., *uchitel'ya* [teacher-ACC], *uchitel'yem* [teacher-INS]) and *k* for feminine nouns (e.g., *polku* [shelf-ACC], *polkoj* [shelf-INS]) whereas other nouns lacked this feature. After brief exposure to a set of inflected nouns, infants showed evidence of learning the underlying gender categories by distinguishing the grammatical forms of nouns, such as the feminine *ruchku* [hand-ACC] or *ruchkoj* [hand-INS], from ungrammatical forms, such as **ruchkyem*. However, in a control condition lacking any form-based gender cues, infants failed to demonstrate learning.

Extending these findings to L2 learning in adults, Kempe and Brooks (2001; Brooks, Kempe, & Donachie, 2011) used the gender-specific diminutive suffixes, –ik (masculine) and –ka (feminine), as form-based cues to Russian gender categories. They observed that these cues facilitated learning of Russian gender-agreement and case-marking patterns and aided learners in generalizing the patterns to new vocabulary.

Second, increasing the number of different lexical items instantiating a grammatical pattern helps learners to extract underlying regularities. Gómez (2002) exposed learners to an artificial grammar that generated non-adjacent dependences between first and last elements (e.g. *jic*, *rud*) surrounding bi-syllabic nonwords (e.g. *kicey, wadim*). Both adults and infants were more successful in learning the non-adjacent dependencies when there were 24 different bi-syllabic nonwords as opposed to just 2 or even 12. Thus, increasing the entropy of the pattern (i.e., the degree of uncertainty about which particular word instantiates the pattern) strengthened the predictive relationship of the crucial non-adjacent dependencies. Gómez's findings echo well with claims (Marchman & Bates, 1994) that learners need to encounter a critical mass of lexical items instantiating a morphological pattern (e.g., English past tense) to generalize it to new items. However, conceiving of a critical mass as an effect of the number of different words instantiating a pattern does not take into account the frequency of occurrences of the individual words—whether different words tend to occur equally often or whether specific words occur more often than others. Having a balanced distribution of words increases the entropy of a grammatical pattern and is thought to promote extension of the pattern to new words (Matthews & Bannard, 2010).

First language (L1) acquisition research indicates that children are more productive with grammatical patterns having a balanced as opposed to a skewed distribution (Matthews & Bannard, 2010). A similar advantage for learning grammatical patterns instantiated by balanced

input has been demonstrated in adult L2 learning (McDonough & Nekrasova-Becker, 2014; McDonough & Trofimovich, 2013; Year & Gordon, 2009). For example, McDonough and Trofimovich exposed participants to the Esperanto transitive construction while varying the distribution of nouns in the construction (balanced vs. skewed to favor specific nouns) and task instructions (inductive vs. deductive). Note that the Esperanto transitive utilizes both subject-verb-object and object-verb-subject word order while marking the object with the accusative suffix –n. Participants in both groups were told that Esperanto word order was flexible; the inductive condition was told to pay attention to the ends of nouns whereas the deductive condition was told that the suffix –n would be added to the noun that functioned as the object. After one session of aural exposure to 24 sentences (12 per word order), only participants exposed to balanced input in the deductive condition were above chance in identifying the object of transitive sentences containing new nouns, which suggested that a balanced distribution of items facilitates learning as long as the underlying rule is made explicit.

The evidence discussed so far suggests that presenting lexical items with similar frequencies in constructions should benefit learning of grammatical categories and associated morpho-syntactic patterns. However, Goldberg and colleagues (Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004; Goldberg, Casenhiser, & White, 2007), noting that all natural languages exhibit skewed word-frequency distributions in accordance with Zipf's law (Mandelbrot, 1953), have made the opposite claim: skewed input should facilitate the learning of grammatical constructions. In their studies, English-speakers were exposed to a novel "appearance" construction (SOV-o as in *The king the ball moopo-ed*) consisting of a nonce verb ending in –o (e.g., *moopo*) used in combination with English nouns; following 16 trials of exposure using five different nonce verbs, participants were tested on their comprehension of the

construction in a forced-choice test. Goldberg and colleagues (2004, 2007) reported that adults were more accurate in identifying the meaning of the construction with skewed input, i.e. when the majority of exposure trials used the same nonce verb, as opposed to a balanced frequency of different verbs. They interpreted the benefit of stacking the input with a single "prototype" as evidence for a meaning-driven process, wherein the argument-structure construction takes on the semantics of the prototypical verb.

Despite certain findings in favor of a skewed input, it remains to be seen whether this is true for inflectional morphology. Thus, in the current study, we tested whether the distribution of items in the input would affect learning of a miniature Russian case-marking paradigm. We did not expect an advantage for skewed input to extend to the realm of inflectional morphology because a noun's lexical-semantic representation cannot serve as a prototype for the different thematic roles expressed in its case-marking paradigm. Even though different noun classes (e.g., masculine and feminine nouns) take different case-marking inflections, the inflections of the different noun classes express the same set of thematic roles. Alternatively, we hypothesized that input containing a balanced distribution of diverse lexical items might facilitate L2 learning and generalization of case-marking inflections, relative to a skewed distribution, by giving the learner greater opportunity to discover the scope of application of each inflection.

Cognitive abilities that affect L2 morphology learning

Input manipulations, like the ones described above, may differ in their effects on the performance of individual learners who bring different sets of strategies and abilities to the task of learning a foreign language, yielding so-called aptitude-by-treatment interactions (McLaughlin, 1980; Robinson, 2001). Indeed, when Brooks, Kempe, and Sionov (2006) examined whether increasing the number of different lexical items in the input, while holding constant the total

number of learning trials, would promote generalization of Russian case-marking patterns, they found an interaction between the input manipulation and participants' scores on the Culture-Fair intelligence test (Cattell & Cattell, 1973), a measure of *g* or general nonverbal intelligence (Duncan et al., 2000), which is presumed to be an index of foreign language-learning aptitude (Grigorenko et al., 2000). Only participants with above-average nonverbal intelligence showed the predicted benefit from more variable input in generalizing morphological patterns to new nouns. For participants with below-average nonverbal intelligence, keeping track of a greater number of nouns in the input appeared to interfere with learning. In the current study, we evaluated possible aptitude-by-treatment interactions by exploring whether effects of varying the distributional characteristics of the input differed depending on participants' nonverbal intelligence.

We also considered the impact of individual differences in statistical learning, defined as "the detection of patterns of covariation between elements in complex stimulus domains" (Reber, 2015, pp. viii). Statistical learning appears to be uncorrelated with measures of intelligence (Gebauer & Mackintosh, 2007; Reber, Walkenfeld, & Hernstadt 1991; Siegelman & Frost, 2015; but see Robinson, 2005, for a negative correlation), yet displays considerable variability that has been linked to individual differences in L1 development in children (Kidd, 2012; Lum, Conti-Ramsden, Morgan, & Ullman, 2014) as well as L1 sentence processing in adults (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak, Christiansen, & Tomblin, 2010; Misyak & Christiansen, 2012). To date, only a few studies have examined whether statistical learning might be linked to individual differences in foreign-language acquisition (Frost, Siegelman, Narkiss, & Afek, 2013; Granena, 2013; Kaufman, DeYoung, Gray, Jiménez, Brown & Mackintosh, 2010, Linck et al., 2013). Kaufman and colleagues (2000) linked British secondary students' performance on the alternating serial reaction time (SRT) task, which measures the extent to which

participants come to anticipate the locations of successive elements of a probabilistic sequence appearing in an array, with their foreign-language proficiency exam scores. Granena (2013), using the same SRT task, demonstrated associations between learning of such sequential patterns and late bilinguals' sensitivity to L2 (Spanish) grammatical rules. Frost and colleagues (2013) linked performance on a visual statistical learning task, involving the extraction of three-element triplets from a continuous sequence, with L2 (Hebrew) literacy acquisition. Two other studies have linked performance on statistical-learning tasks to learning of syntactic structures in artificial or semi-artificial languages (Hamrick, 2015; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014).

To gain a better understanding of what aspects of morphology learning are predicted by statistical-learning ability, Brooks & Kempe (2013) exposed naïve English-speaking participants to a miniature version of Russian over six language-learning sessions. Statistical-learning ability was assessed using an auditory artificial grammar learning (AGL) task shown to correlate with L1 sentence processing in adults (Misyak & Christiansen, 2012), wherein participants listened to "sentences" consisting of nonwords arranged in sequences specified by an artificial grammar and were tested on their ability to distinguish sequences that obeyed the grammar from those that did not. Although statistical-learning ability was predictive of participants' accuracy in inflecting the nouns in the training set (i.e., item-based learning of case-marking suffixes), it failed to predict generalization of the Russian morphological patterns to new vocabulary, which instead was related to nonverbal intelligence. Statistical-learning ability was also unrelated to vocabulary recall; instead, vocabulary recall was predicted by measures of phonological short-term memory (nonword repetition) and verbal working memory (reading span). So far, the results linking statistical learning with L2 morphology learning and temporary-storage capacity with L2

vocabulary recall might be taken as support for Ullman's (2001, 2004) proposal that the implicit procedural-memory system (underlying statistical learning as well as motor and cognitive skill acquisition) plays a critical role in grammar learning and a lesser role in vocabulary acquisition. However, when participants were asked at the end of the experiment what structural regularities they had noticed, their degree of awareness of the underlying rules mediated the effect of statistical learning and nonverbal intelligence on morphology learning, but not the effect of short-term memory capacity on vocabulary learning. This finding challenges the idea that morphology learning is exclusively underpinned by implicit procedural knowledge, but is in line with recent findings on the role of awareness in associative learning (Weidemann, Satkunarajah, & Lovibond, 2016).

Counter to the view that separable mechanisms underlie vocabulary and grammar acquisition, researchers (Erickson & Thiessen, 2015; Saffran, Aslin, & Newport, 1996) have conceptualized statistical learning as a mechanism for extracting recurring "chunks" comprising adjacent syllables in continuous speech, with such chunks serving as input for vocabulary learning. Supporting this hypothesis are reports of correlations between statistical learning and L1 vocabulary size in children (Evans, Saffran, & Robe-Torres, 2009) and L2 vocabulary learning in adults (Speciale et al., 2004). Likewise, evidence of dependencies between lexical and morphological development (Bates & Goodman, 1997; Marchman & Bates, 1994) suggests that these two domains share underlying learning mechanisms. To address this controversy, the current study re-examines the extent to which statistical-learning ability and short-term and verbal working memory capacity predict morphology and vocabulary learning, and whether explicit awareness of the underlying regularities is required. Building on our previous work on individual differences in L2 learning of Russian case and gender (Brooks & Kempe, 2013; Kempe & Brooks,

2008), the current study examines learning of a more complex case-marking paradigm in which distributional information provides the only available cues to the underlying gender categories. We address the following research questions and hypotheses:

1. Do distributional characteristics of the input impact L2 learning of morphology? Based on previous research (Matthews & Bannard, 2010; McDonough & Trofimovich, 2013) suggesting that high entropy, less predictable input facilitates learning of grammatical patterns, we predicted that, when controlling for individual differences in intelligence, statistical learning, and memory capacity, as well as for prior exposure to other languages (cf. Odlin, 1989), accuracy in producing Russian case-marking inflections would be higher with balanced than with skewed input.

2. Does the effect of the input manipulation differ as a function of participants' cognitive skills? Based on our previous findings (Brooks & Kempe, 2013; Kempe & Brooks, 2008) we hypothesized that nonverbal intelligence would predict L2 morphology learning. However, given that only individuals with higher nonverbal intelligence benefitted from input comprising a larger vocabulary set when learning Russian morphology (Brooks et al., 2006), we predicted that benefits from a balanced distribution might be greater for learners with higher nonverbal intelligence.

3. Is there dissociation between predictors of morphology and vocabulary learning? In line with Ullman's (2004) proposal, we hypothesized that statistical-learning ability, a form of procedural learning, would predict learning of L2 morphology whereas measures of short-term and verbal working memory capacity would predict learning of L2 vocabulary.

Method

Participants

Fifty-four students (28 men, 26 women, mean age 22 years; range 18–33 years) were recruited from a diverse student population at an urban public university in the United States using flyers. Four additional students completed cognitive assessments, but discontinued participation prior to the language-learning sessions. Participants were paid $100 as compensation for their time. As required by the university Institutional Review Board, participation was open to any student, with the exception of those with prior knowledge of Russian or another Slavic language. Most participants (87%) were undergraduates; the remainder comprised graduate students and recent graduates. Participants had studied a variety of disciplines (37% psychology), with only one reporting coursework in linguistics (English major).

Participants completed a foreign-language background questionnaire on which they listed languages spoken at home or studied in school and rated their proficiency in each language (including English) in the domains of reading, writing, listening, and speaking, using a scale of 1 (very poor) to 6 (excellent). All participants had prior exposure to one or more foreign languages (M=2.09, SD=.94). Most common were Spanish (N=36), Italian (N=11), Chinese (N=11), French (N=9), Japanese (N=4), Arabic (N=4), and Urdu (N=3); other languages were American Sign Language, Bamileke, Creole, Filipino/Tagalog, Gujarati, Hebrew, Korean, Latin, Telugu, Vietnamese, Wolof, and Yoruba. On average, participants indicated high proficiency in English (M=5.70, SD=.54), moderate proficiency in an L2 (M=3.16, SD=1.54), and low proficiency in an L3 (M=1.54, SD=1.43). Fewer than half (44%) indicated high proficiency in an L2, indicated by mean self-ratings of 3.5 or higher.

Participants were randomly assigned to input conditions (balanced/skewed), with equal numbers of men and women per condition. Groups assigned to input conditions did not differ in

age, gender, or foreign-language background (self-rated proficiency in an L2 or L3, total number of languages learned, specific languages studied).

Cognitive skills assessments

Nonword-repetition task. To assess phonological short-term memory, we used an established nonword-repetition task (Gupta, 2003; Experiment 2), implemented on a Dell PC using E-prime software (Schneider, Eschman & Zuccolotto, 2002) and audio-recorded in its entirety. Ninety nonwords, recorded by a female native-speaker of English, were assembled into five blocks of 18 trials, with two-, four-, and seven-syllable items (e.g., *chentok, fiscolumbate, meskolonaymendalic*) presented in randomized order (six items per length). On each trial, participants heard a nonword through speakers and were instructed to repeat it as soon as a fixation cross appeared on the computer screen, 100ms after the offset of the nonword. An experimenter sat behind the participant and scored each trial using a binary criterion (correct/incorrect). Score sheets were subsequently checked for accuracy by a trained coder using the audio recording (coding agreement=95.5%). We used the number of correct nonword repetitions (out of 90) as a measure of phonological short-term memory.

Reading-span task. To assess verbal working memory capacity, we used the reading-span task of Daneman and Carpenter (1980). This task has both storage and processing components required of complex span measures shown to correlate with L1 language comprehension (cf. Daneman & Merikle [1996] for a meta-analysis). Participants were instructed to read aloud sets of unrelated sentences while holding in mind the last word of each sentence for future recall. Sentences were presented in sets of increasing size, ranging from two to five sentences per set. Each sentence was printed individually on an index card, with a second card used to reveal each successive word as the participant read the sentence aloud. After each set, participants were shown

a cue to recall the final word of each sentence in the set. The test consisted of 70 sentences, with 5 sets at each set-size. Following procedures outlined by Shah and Miyake (1996), we calculated the reading-span score as the total number of words recalled correctly out of 70 to avoid a restricted range of scores.

Auditory AGL task. To assess statistical-learning ability, we used the 'adjacent-dependencies' AGL task of Misyak and Christiansen (2012), run on a Dell PC using E-prime software. Auditory sequences of nonwords were generated using the following production rules:

S -> NP VP

NP -> d N

NP -> D A N

VP -> V (NP)

The sequences utilized ten nonwords (*hep, tam, biv, dupp, jux, lum, meep, sig, zoet, rauk*) distributed over categories (3 N, 3 V, 2 A, 1 d, and 1 D), with assignment of nonwords to categories randomized for each subject. Participants listened to nonword sequences for approximately 30 minutes (60 sequences presented 3 times each). Immediately after exposure to the input, participants completed 40 two-alternative, forced-choice trials in which they heard two sequences and had to decide which sequence followed the same rules as the input. Half of the grammatical sequences were taken from the 60 sequences presented as input; the other half comprised novel sequences generated by the same rules. We used the total number of correct responses (out of 40) as a measure of statistical learning.

Culture-Fair intelligence test. To obtain an estimate of nonverbal intelligence, we administered the Culture-Fair intelligence test, Scale 3, Form A (Cattell & Cattell, 1973). Participants received a booklet with four sets of multiple-choice problems and an answer sheet.

Each set started with several example problems, followed by 10–14 problems of increasing difficulty. Participants were instructed to solve as many problems as they could in the allotted time (ranging from 2.5–4 minutes per set). Two sets ('Series' and 'Matrices') involved selecting an abstract geometric pattern from six alternatives to complete a series or matrix. One set ('Classification') required participants to identify which two of five patterns were alike (i.e., different from the other three). The last set ('Conditions/Topology') required participants to select a stimulus out of five alternatives that matched another stimulus with respect to the placement of a dot among geometric forms. The test was scored using the provided template. The number of correct items (out of 50) served as a measure of nonverbal intelligence.

Russian language-learning materials

      Vocabulary. 34 Russian nouns served as vocabulary; 9 masculine and 9 feminine nouns were used in the training sessions, and 8 masculine and 8 feminine nouns were reserved to test for generalization of case-marking inflections to new vocabulary (see Appendix). All nouns were bi-syllabic to minimize variation in ease of pronunciation across noun genders (i.e., it is possible to construct a set of monosyllabic masculine nouns, but not feminine nouns). Within each gender, eight nouns were stressed on the first syllable and nine on the second syllable; including both stress patterns was necessary to generate sufficient items for training and test sessions. Lexical stress was counterbalanced across training and testing items and across genders to maximize the phonological heterogeneity of items within each gender.

      Pictures and dialogs for case-marking elicitation. Each noun was associated with a corresponding line drawing taken from the Snodgrass and Vanderwart (1980) set of standardized pictures. The line drawing of each object was presented on its own (for nominative case) and in

conjunction with a picture of an elephant moving towards (for dative case), from (for genitive case), or under the object (for instrumental case); see Figure 1 for an example set.

Participants learned the Russian words and phrases through exposure to dialogs created for each semantic context; see Table 1 for examples of phrases containing masculine and feminine nouns. Note the use of specific Russian prepositions for each semantic context: *k* [towards] for dative, *ot* [from] for genitive, and *pod* [under] for instrumental. Within each gender category, all nouns used the same allomorphs; thus, all feminine nouns took the dative suffix –*e*, the genitive –*y*, and the instrumental –*oj*, and all masculine nouns took the dative –*u*, the genitive –*a*, and the instrumental –*om*.

Input conditions. Two input conditions varied the distribution of nouns across the 108 trials of each language-learning task (described below): In the balanced condition, 18 nouns each occurred six times per task (two dative, two genitive, and two instrumental trials). In the skewed condition, two nouns occurred 18 times per task (six dative, six genitive, six instrumental), two nouns occurred 12 times per task (four dative, four genitive, four instrumental), two nouns occurred six times per task (two dative, two genitive, two instrumental), and 12 nouns occurred three times per task (one dative, one genitive, one instrumental). See the Appendix for the frequencies of each noun in the balanced and skewed conditions.

Procedure

Participants were told at the outset that they would be learning Russian. Throughout the study, they were exposed to Russian in spoken phrases paired with corresponding pictures, without English translations or Russian written forms. Aural exposure to the Russian circumvented the need for participants to learn the Cyrillic alphabet or a transliteration scheme for Russian orthography. Use of a transliteration scheme potentially would have made the task more difficult

due to interference from English (e.g., the Russian dative suffixes *–e* and *–u* are not English vowels).

Participants were tested individually and completed three computer-based language-learning sessions lasting approximately 2-hr each. Sessions were scheduled approximately one week apart (M=6.7 days; SD=5.3), but varied due to missed appointments and exam schedules. Intervals between sessions did not differ as a function of the input condition (*p*=.56). Session 1 consisted of training only, and sessions 2 and 3 consisted of training followed by testing. Testing was not included in session 1 due to time constraints imposed by the need to go over instructions before each task, and participants' slow rate of responding due to their lack of familiarity with the Russian dialogs. Questionnaires and assessment tests were administered in a separate session (approximately 2hr duration), scheduled at participants' convenience prior to the language-learning sessions. The language background questionnaire was administered first, followed by the four assessment tests in a randomized order.

Task presentation was controlled using PsyScope experimental software (Cohen, MacWhinney, Flatt, & Provost, 1993) run on an Apple computer. All Russian phrases were presented through headphones, with corresponding pictures appearing on the computer screen. A research assistant manually advanced the trials to allow participants as much time as needed to respond. After each trial, participants received computer-generated feedback: the program repeated the correct Russian phrase while showing the corresponding picture. The research assistant did not speak any Russian and provided only non-specific encouragement ("You are doing fine"). Due to the complexity of the procedure involving different language-learning tasks, research assistants were instructed to answer procedural questions after the participant had read

the task instructions, but not to provide translations of Russian words or phrases. Sessions were audio-recorded to ensure that procedures were followed.

Training tasks. Training consisted of four language-learning tasks (15–20min duration per task) presented in a fixed order: listen and repeat, noun comprehension, case comprehension, and production, see Figure 2 for a depiction of a trial for each task. Each task comprised 108 trials (36 dative, 36 genitive, and 36 instrumental) presented in random order.

The listen-and-repeat task provided exposure to the Russian dialogs and their corresponding pictures (i.e., semantics), while requiring participants to practice producing the Russian phrases through imitation (i.e., perception and production of unfamiliar sequences of sounds). Participants viewed a series of pictures one-at-a-time while listening to short dialogs that described each picture. They were shown the instructions: *You will see a picture. At the same time, you will hear a little dialog in Russian. In this dialog, the male speaker asks a question, and the female speaker answers his question. Your task is to repeat the answer. In other words, always repeat what the female speaker says. After you have repeated the answer, the experimenter will press a key and then you will hear the answer again. Please repeat it a second time. After that the experimenter will press a key to show you the next picture. Please ask the experimenter if you have any questions. If not, the experimenter shall press a key, and the first picture will appear. Have fun!* Thus, as illustrated in Figure 2, when shown a picture of an elephant under a star, the participant heard the male speaker ask the question *Gde pryachetsya slon?* [Where is elephant hiding?], followed by the female speaker's answer *pod zvezdoj* [under star]. After the participant repeated the phrase *pod zvezdoj*, the phrase was presented again with the participant instructed to repeat it a second time.

The noun-comprehension task prompted participants to listen to the dialogs to identify the nouns in the phrases. Participants were shown the following instructions: *Now you will see three pictures on the screen. Again, you will hear the little dialog. Listen carefully to both the question and the answer, and then decide which object was mentioned in the dialog. Select the appropriate picture by pressing the corresponding button on the button box. If the female speaker mentioned the object in the left picture, press the LEFT (red) button. If the female speaker mentioned the object in the middle picture, press the MIDDLE (yellow) button. If the female speaker mentioned the object in the right picture, press the RIGHT (blue) button. After you have pressed the button, you will see the correct picture displayed in the middle of the screen. Please repeat the corresponding phrase as you see the correct picture. The experimenter will press a key when you are ready to see the next set of pictures.* Thus, as illustrated in Figure 2, each trial depicted three different objects in the same semantic context, with all three nouns of the same gender, e.g., *zvezda* [star], *shlyapa* [hat], *plita* [stove]. Upon hearing each dialog, participants used a three-option button box to make their choices, with the positions of the different objects in each triad held constant across trials.

In the case-comprehension task, participants heard the same dialogs and were tested on their comprehension of the Russian prepositions for dative, genitive, and instrumental cases. They were shown the instructions: *In the next part, you will again see three pictures. You will also hear a dialog. Listen carefully to find out which picture matches what you hear, and then select the appropriate picture by pressing the closer button. In this test, there will be the same object in each picture. In the left picture the elephant moves away from the object. In the middle picture, the elephant is under the object, and in the right picture, the elephant moves toward the object. If the dialog describes the elephant moving away from the object, then press the LEFT button. If the*

dialog describes the elephant under the object then press the MIDDLE button. If the dialog describes the elephant moving towards the object then press the RIGHT button. After your choice, you will see the correct picture in the middle of the screen. Please repeat the corresponding phrase as you see the correct picture. The experimenter will press a key when you are ready to see the next set of pictures. As illustrated in Figure 2, the same object appeared in each semantic context. Upon hearing each dialog, participants used a three-option button box to make their choices, with the positions of three semantic contexts held constant, as indicated in the task instructions.

In the production task, participants' mastery of the case-marking inflections was probed by requiring them to produce the female speaker's answers from the previous dialogs. Participants were shown the instructions: *This is the last part. You will now see a picture of one of the objects. The female speaker will describe it by using a simple Russian sentence that can be translated as "This is [OBJECT]." Listen to this sentence carefully and try to remember the name of the object. Next, you will see a picture, and you will hear the male speaker asking a question. Please answer this question in Russian (remembering phrases that the female voice said in the previous dialogs). Speak clearly into the microphone. After that, the experimenter will press a key, and you will hear the correct answer. Please repeat the correct phrase. Try to speak clearly and remember how to answer the male speaker's questions. The experimenter will press a key for you to see the next set of pictures.* As illustrated in Figure 2, the first part of the trial reminded participants of the noun to be used in the required answer, with the noun introduced in nominative case with its corresponding line drawing. The second part of the trial required participants to produce the appropriate preposition based on the semantic context, i.e., *k* [towards] for dative trials, *ot* [from] for genitive trials, and *pod* [under] for instrumental trials, in combination with the case-inflected noun.

Testing tasks. Testing comprised two tasks of approximately 25min duration. Testing was conducted on the same days as training sessions 2 and 3, with an approximately 10min break between training and testing. Tests were audio-recorded and transcribed in their entirety.

The production test was identical in format to the production task used in training, except that new vocabulary items were randomly interspersed to examine participants' ability to generalize case-marking patterns beyond the trained items. Participants were given written instructions: *This part will be very similar to the last part of the training. You will see a picture of one of the objects. The female speaker will describe it by using a simple Russian sentence that can be translated as "This is [OBJECT]." Most of the objects will be familiar. You have practiced those in the previous sessions. However, some of the objects will be new. Listen to the sentence carefully and try to remember the name of the object. Next, you will see a picture, and you will hear the male speaker asking a question. Please answer this question in Russian. Speak clearly into the microphone. After that, the experimenter will press a key, and you will hear the correct answer. Please repeat the correct phrase. Remember to speak clearly. The experimenter will press a key for you to see the next set of pictures.* Note that each test presented 54 'old' and 24 'new' trials (18 old and 8 new nouns appearing once per semantic context), with different new nouns used in each test session.

After the production test, participants completed a vocabulary test to determine whether they could recall the names of the 18 objects used in training. Participants read the instructions: *You will now see all the pictures one at a time. Look carefully at the object, and try to recall the Russian name of it. Just say the name, not any of the other words from the phrases. If you can recall it, even vaguely, say it into the microphone. It's OK if the word does not sound perfect. Try*

to do as good a job as you can. If you cannot recall it, say "don't know" into the microphone. When you are done speaking, the experimenter will press a key to show the next picture.

Exit questionnaire. To assess awareness of the underlying rules, participants completed a brief written exit questionnaire following session 3. Participants were asked (1) whether they noticed any patterns in the Russian words and phrases and to describe the patterns, (2) to describe their strategy for answering the questions about the position of the elephant (towards, from, or under the object), (3) whether they noticed anything about the new vocabulary that helped them to use the words in phrases and to describe what they noticed, and (4) to describe what they had discovered about the structure of Russian words, phrases, and grammatical patterns. We used the brief questionnaire rather than an elaborate think-aloud protocol (cf. Leow & Morgan-Short, 2004) because implementation of the latter was not practical with our tightly controlled methodology.

Coding of the language-learning tasks

To assess comprehension of the question-answer dialogs and the corresponding prepositions used for dative, genitive, and instrumental case, we assessed performance on the case comprehension task in each training session. Accuracy was determined from the computer log, which recorded whether the correct picture was selected for each trial.

Responses to the production and vocabulary tests were transcribed and coded by a native speaker of Russian (VK). Case production was coded as correct if the participant provided the correct preposition and suffix on the target noun, regardless of stress. Vocabulary recall was coded as correct if the initial phoneme as well as the basic form of the word stem resembled the target noun, irrespective of the final vowel. That is, any errors in inflecting the noun were not counted as vocabulary errors. To ascertain the consistency of the coding, 15% of the data were re-transcribed and re-coded by the same native speaker without access to the previous transcription (cf. Gwet,

2014; Polio & Shea, 2014). Intra-rater reliability (percentage agreement between the rounds of coding) was very high with 94.4% agreement for case production and 98.3% for vocabulary recall.

Coding of the exit questionnaire followed a simple scheme that did not distinguish between different levels of awareness or depth of processing as is customary in the analyses of think-aloud protocols (Leow, Grey, Marijuan & Moorman, 2014). The simplified coding was justified by the fact that participants' responses provided limited data density that did not lend itself to more fine-grained analysis. Our main interest was to evaluate whether learners had noticed any patterns at all and whether the noticed regularities pertained to the case inflections or gender categories. Participants received a point for case if they mentioned that Russian nouns had "suffixes", "endings", or "conjugations" that changed depending on the elephant's position (semantic context). Participants received a point for gender by mentioning that different words took different sets of endings or that there was more than one ending for each position, listing at least one pair of endings that expressed the same meaning for different sets of words, or mentioning that the endings depended on the basic form of the noun (vowel or consonant ending). Thus, scores ranged from 0 (no awareness) to 2 (awareness of both case and gender). Two coders independently scored the exit questionnaires, with 98.3% agreement on case and 96.7% agreement on gender.

Results

We first present descriptive statistics and correlations between predictor variables and outcome measures. Case comprehension data were obtained from all three training sessions. Case production for old and new items and vocabulary recall data were obtained from the tests that followed training sessions 2 and 3. Awareness was assessed via the exit questionnaire following session 3. Table 2 shows the means, standard deviations and ranges for the predictor variables as a function of input condition. Table 3 presents the language-learning outcomes as a function of

input condition and session. None of the predictor variables or outcome measures showed significant group differences between input conditions (negative values of the *t* statistic indicate numerically greater values for the skewed than for the balanced condition). Note that overall accuracy in case production for old items did not differ between input conditions; however, for the skewed condition, accuracy was higher for high-frequency nouns presented 18 times (session 2: 56.8%, session 3: 76.5%) than for low-frequency nouns presented three times (session 2: 45.7%, session 3: 60.2%).

Table 4 presents zero-order correlations between scores on the cognitive skills assessments, collapsed across input conditions. None were significant after Bonferroni correction (α=.005), although the direction of effects was in line with previous findings indicating weak (positive) correlations between nonword repetition, reading span, Culture-Fair intelligence, and auditory AGL tasks (Brooks & Kempe, 2013).

Comparison of language-learning outcome measures

To compare performance across the different outcome measures while exploring the temporal dynamics of performance, we performed an ANOVA, with participants as random effects, on the percentages of correct responses. Outcome measure (case comprehension, case production for old items, case production for new items, vocabulary recall) and session (2 vs. 3) were within-participant variables and input condition was a between-participants variable in the ANOVA. Note that the purpose of this analysis was to compare the outcome measures over time and not to examine the effect of the input manipulation, which requires inclusion of learner-based predictor variables into the model. We applied Bonferroni corrections to all post-hoc comparisons.

The analysis yielded main effects of outcome measure, $F(3,50)=358.4$, $p<.001$, and session, $F(1,52)=87.6$, $p<.001$, and an interaction between outcome measure and session,

$F(3,50)=4.9$, $p<.001$. Post-hoc t-tests revealed that all tasks differed significantly from each other (all p's$<.001$), with highest accuracy for case comprehension, followed by vocabulary recall, case production for old items, and case production for new items. Although performance improved from session 2 to 3, gains in performance for case comprehension were smaller than for other outcome measures, due to near-ceiling performance on case comprehension in session 3. Because case comprehension was also assessed in session 1, we performed an additional ANOVA on case comprehension accuracy with session (1, 2, 3) as a within-participants variable and input condition as between-participants variable. This analysis showed a main effect of session, $F(2,51)=26.0$, $p<.001$, with post-hoc t-tests showing significant differences between all three sessions, indicative of continuous improvement.

Clustering of language-learning outcomes

Table 5 presents zero-order correlations between the outcome measures. To ascertain whether separate analyses for comprehension vs. production data were justified (i.e., whether outcome measures clustered by type of task or by session), we performed a Principal Component Analysis using Varimax rotation with Kaiser normalization of the ten outcome measures, with participants as a random effect. The analysis yielded three factors with Eigenvalues greater than 1, see Table 6 for factor loadings. The first factor, accounting for 50.3% of the variance linked awareness with case production for old and new items. The second factor, accounting for 15.1% of the variance, linked case comprehension with case production for old (both sessions) and new items (only in session 2). The third factor, accounting for 11.5% of variance, was associated with vocabulary recall. These results indicate clustering by task type rather than session and justify separate analyses of comprehension, production, and vocabulary recall.

Modeling effects of predictor variables on language-learning outcomes

Below, we report analyses designed to test our hypotheses. Because working memory capacity may be moderately correlated with nonverbal intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002) and statistical learning (Misyak & Christiansen, 2012; but see Siegelman & Frost, 2015), we simultaneously entered all of the cognitive measures as predictors into logistic regression models. To account for variability introduced by the different nouns as well as participants, we fitted mixed-effects logistic regression models with a logit link function and crossed random effects (Jaeger, 2008) to each outcome measure (case comprehension, case production for old items, case production for new items, vocabulary recall) in each session, with input condition coded as a dummy variable. The models included random intercepts of nouns and participants, and random slopes of input condition on nouns. The predictor variables were scores on the cognitive skills assessments: auditory AGL, Culture-Fair intelligence, nonword repetition, reading span, and the total number of languages learned.[1] As nonword repetition, reading span, and total number of languages learned failed to yield any significant effects on outcomes, they were excluded from the final models shown in Table 7.

First, we hypothesized that a balanced distribution of noun tokens would lead to superior learning of noun morphology compared to a skewed distribution. Recall that we did not find any group differences between the two input conditions, as indicated in Table 3. However, with the model controlling for individual differences in auditory AGL and nonverbal intelligence, we found that the input condition yielded a significant effect on generalization performance during session 2, such that for novel nouns there was a 42% increase in the odds of producing the correct case marker in the balanced condition compared to the skewed condition when nonverbal intelligence and AGL were controlled (see Table 7). However, this effect was not evident at session 3, which suggests that the potential benefit of less predictable input diminished over the course of learning.

Second, we hypothesized that nonverbal intelligence would be linked to morphology learning. The analyses showed that Culture-Fair intelligence predicted case comprehension in the initial session, and case production for old and new items in the later sessions. In contrast to previous findings (Kempe & Brooks, 2008) that linked Culture-Fair intelligence with vocabulary recall—presumably because more efficient learning of morphological patterns aided vocabulary retention—it did not predict vocabulary learning in this study. To test for an aptitude-treatment interaction between nonverbal intelligence and the input manipulation, we included the corresponding interaction term in the logit mixed-effects logistic regression models for each outcome variable. The only marginally significant interaction involved case comprehension at session 3 ($z=2.04$, $p<.05$). To interpret the interaction, we computed correlations for each input condition separately. The correlation between case comprehension and nonverbal intelligence was significant in the skewed, $r=.44$ ($n=27$), $p<.05$, but not the balanced condition, $r=.10$ ($n=27$), $p=.61$. One might interpret this interaction as suggesting that, at least for some aspects of L2 morphology learning, nonverbal intelligence might play a greater role in conditions where the distributional characteristics of the input are suboptimal. However, given that the participants were near ceiling in case comprehension in session 3 (averaging 95.3%), this interpretation may not be warranted.

Third, we examined whether individual differences in statistical-learning ability might differentially predict morphology and vocabulary learning. We found the auditory AGL task to predict accuracy in case comprehension across sessions, and accuracy in case production for old items in session 2, with the odds of identifying or producing a correct case marker increasing between .3% and 1.2% for each percent increase in correct performance on the AGL task, all other factors being equal. In contrast, auditory AGL was unrelated to accuracy in generalizing case-marking inflections to new vocabulary or in vocabulary retention. Additionally, neither nonword

repetition nor reading span yielded any significant effects in the regression models, thus failing to provide evidence for a link between short-term or verbal working memory capacity and vocabulary learning.

We also sought to explore whether explicit awareness of case and gender mediated effects of nonverbal intelligence and statistical-learning ability on learning outcomes. To determine what factors predicted awareness, we performed a multiple regression analysis with awareness as the dependent variable, with input condition, Culture-Fair intelligence, auditory AGL, and total number of languages learned as predictors, and participants as random effects. This analysis indicated that total number of languages learned (β=.23, t=2.43, p<.05) and Culture Fair intelligence (β=.04, t=1.96, p=.05) predicted the likelihood that participants would gain awareness of the features of Russian morphology (case and gender) over the course of the study, with the model accounting for 21% of the variance, $F(4,48)$=3.20, p<.05. We then ran a series of mediation analyses employing bootstrapping to estimate the 95% confidence intervals for the indirect effect of nonverbal intelligence on learning outcomes (mediated by awareness); mediation is deemed statistically significant at p=.05 if the confidence interval does not include zero. These analyses indicated that awareness mediated effects of Culture Fair intelligence (as well as total number of languages learned) on case production for old and new items in session 2, and case production for new items in session 3 (confidence intervals ranged from .0001 to .07). Note, however, that the indirect effect of total number of languages learned on the learning outcomes occurred in the absence of a significant direct effect.

Discussion

Using a miniature Russian case-marking system consisting of masculine and feminine nouns in nominative, dative, genitive, and instrumental case, we tested whether manipulating the

distribution of nouns instantiating the morphological patterns would impact learning outcomes, while also examining individual differences in cognitive abilities relevant to L2 learning. We found evidence that balanced, higher entropy input was beneficial for generalizing case inflections to new vocabulary, although the effect of the input manipulation was weak and transient and did not interact with any of the individual differences measured. We found that statistical-learning ability was predictive of learning to inflect nouns in the training set, whereas nonverbal intelligence predicted both learning and generalization of inflections to new nouns, with the latter largely mediated by participants' explicit awareness. In contrast, none of the predictors, including temporary storage measures, were associated with vocabulary recall.

Effects of the input manipulation

The observed benefit of balanced input, albeit transient, is in line with the view that decreasing the predictability of the input by increasing the variability of lexical items promotes pattern learning (Gómez, 2002; Matthews & Bannard, 2010). While balanced distributions emphasize the diversity of lexical items instantiating grammatical patterns, skewed distributions draw attention to a small number of items at the expense of others, thereby promoting item-based learning and reducing the beneficial effects of high type frequency on generalization. In other words, even if the total number of different words in the input is the same, skewed distributions tend to focus learners' attention on the morphological properties of just a few items, which reduces the set from which the underlying regularities can be extracted. This finding contrasts with the skewed-advantage hypothesis (Casenhiser & Goldberg, 2005; Goldberg et al., 2004) that learners acquire grammatical constructions more readily when the input favors specific items which instantiate prototypical meanings associated with specific verbs. As an example, consider the English double-object construction [Subject V Object$_1$ Object$_2$] as in *Jane gave me the letter*. If a

specific verb, such as *give*, repeatedly occurs in the construction, learners may come to associate the construction as a whole with the verb's semantic representation, involving transfer of something from one person to another. Thus, when encountering a new verb in the construction, e.g., *Karen mooped him the book*, learners may interpret *moop* as indicating a specific method of transfer. Although Goldberg and colleagues (2004, 2007) documented benefits for skewed input with artificial verbs, efforts to apply the skewed-advantage hypothesis to L2 learning have yielded contradictory results. In a study of Thai college students' acquisition of the English double-object construction, McDonough and Nekrasova-Becker (2014) found balanced input to be more successful in promoting comprehension of the construction than skewed input when tested 2 weeks post-exposure. Likewise, in a study of Korean children's acquisition of the English double-object construction, Year and Gordon (2009) found balanced input to be more beneficial in promoting productive usage in a delayed post-test (6 weeks post exposure), with no differences between input conditions observed in immediate post-tests. In an effort to replicate Casenhiser and Goldberg (2005) more directly by using their novel construction with L2 learners, Nakamura (2012) also failed to find an advantage for skewed input in Japanese college students with English proficiency, casting further doubt on the applicability of Goldberg's hypothesis to L2 learning of argument-structure constructions.

We view this hypothesis as even less applicable to L2 learning of inflectional morphology. In case-marking systems like Russian, the meanings of individual suffixes cannot be associated with the lexical semantics of individual nouns because every noun combines with a set of case-marking suffixes. In the current study, the prepositions, rather than the nouns, distinguished the semantic contexts (*towards*, *from*, *under*), thus providing clues to the functions of the suffixes. For learners to extend case marking correctly to new vocabulary requires extraction of form-based

cues to the underlying gender categories. Learners must come to recognize that feminine nouns take different suffixes than masculine nouns, with nominative noun endings providing a cue to noun gender (e.g., –*a* for feminine, consonant for masculine).

In our study, the beneficial effect of a balanced distribution was found for generalization of case-marking to new vocabulary, and did not extend to case comprehension or production of case-marking inflections on the training nouns. During case-comprehension trials, learners had to choose from sets of three pictures depicting the semantic relations associated with the three cases (with all relations depicted with the same object). While the case comprehension task provided learners with additional exposure to the inflected forms, it did not require participants to map each suffix onto a specific case or categorize nouns based on their morpho-phonological features. Successful case comprehension could be achieved based on understanding the questions (i.e., with distinct verbs in each question providing salient cues to distinguish them) and/or the Russian prepositions (i.e., *k* [towards], *ot* [from], *pod* [under]). The lack of an effect of the input manipulation on case comprehension suggests that learning the mappings between the semantic contexts, questions, and prepositions did not depend on the distribution of nouns instantiating the patterns.

The differential effects of the input manipulation on case production for old and new items, in combination with higher case production accuracy for old than for new items across input conditions, suggest that participants may have relied on different strategies for the two types of items. For old nouns presented in the training set, participants could select the correct suffixes by relying on item-based learning (i.e., associating suffixes with individual nouns) or category-based learning (i.e., associating sets of suffixes with gender categories). In contrast, to generalize case-marking, learners had to categorize the new nouns to select the appropriate suffixes based on their

forms in the nominative case in phrases like *Eto kofta* [this is blouse] or *Eto arbuz* [this is watermelon]. Higher accuracy in case production for old than for new items suggests that item-based learning boosted performance for old items, with a similar effect across input conditions. Higher accuracy in generalization suffixes to new items with balanced as opposed to skewed input suggests that the input manipulation affected category-based learning.

The effect of the input manipulation, however, was short-lived and did not extend to session 3, wherein we probed for participants' awareness of the features of the Russian case-marking system and found no differences between input conditions. The fact that the benefit of balanced input was transient is a novel finding, revealed by our extended training regimen and multiple test sessions that contrast with procedures examining L2 learning of inflectional morphemes after a single session (e.g., McDonough & Trofimovich, 2013, who examined L2 acquisition of an accusative suffix in the Esperanto transitive construction). Our results suggest that over time, participants exposed to skewed input accumulated sufficient exposure to the less frequently occurring lexical items to extract the underlying regularities, even if the skewed input initially focused their attention on a limited set of items. Given that all languages exhibit skewed word-frequency distributions (Mandelbrot, 1953), the learning mechanisms that underlie the detection of grammatical patterns must be fairly robust to variation in word frequencies.

Aptitude-by-treatment interactions

Given that learners with varying cognitive aptitudes may respond differently to an input manipulation (Robinson, 2001), we explored statistical models that included aptitude-by-treatment interactions. The results failed to provide compelling evidence of such interactions, as the only effect to reach statistical significance involved performance on case comprehension in session 3, where accuracy approached ceiling. The lack of aptitude-by-treatment interactions involving

nonverbal intelligence might seem to be at odds with our prior findings (Brooks et al., 2006) that the benefits of more variable input were evident only for participants with above-median nonverbal intelligence. Whereas the prior study manipulated the quantity of different nouns instantiating a case-marking paradigm, the current study introduced a more subtle manipulation of the token frequency by distributing the same number of nouns differently across trials. Our results suggest that the transient benefit of a balanced distribution of noun tokens was independent of individual differences in nonverbal intelligence.

Individual predictors of morphology learning

To gain further understanding of the mechanisms underlying L2 morphology and vocabulary learning, we used a principal component analysis to identify patterns amongst the outcome measures collected over three sessions. This analysis revealed three components: The first strongly linked participants' explicit awareness of Russian case and gender with case production for old and new items; the second linked case comprehension with case production independently of awareness, and the third was strongly associated with vocabulary recall. This pattern is inconsistent with models emphasizing continuities in lexical and grammatical development (Bates & Goodman, 1997; Marchman & Bates, 1994). Instead, the dissociation in performance in these two domains seems more consistent with the idea that different mechanisms underlie morphology and vocabulary learning (Ullman, 2004).

Our finding that performance on the auditory AGL task, a measure of statistical learning, predicted L2 learning of inflectional morphology, but not vocabulary recall, provides further evidence for this conclusion. Notably, however, this finding seems to be at odds with Speciale and colleagues (2004) who reported correlations between L2 vocabulary acquisition and another statistical learning task, phonological sequence learning, involving recognition of nonwords

recurring in an auditory sequence. Further research is necessary to determine whether different measures of statistical learning index the same underlying construct. Although performance on the AGL and phonological sequence learning tasks has not been compared in the same set of participants, two studies (Gebauer & Mackintosh, 2007; Siegelman & Frost, 2015) comparing different measures of statistical learning (i.e., serial reaction time, AGL, and/or speech-stream tasks) failed to yield evidence of a unitary construct. This leaves open the possibility that individual differences in vocabulary learning might be linked with aspects of statistical learning that are not captured by the AGL task.

In contrast to a previous study (Brooks & Kempe, 2013), we failed to find significant relationships between temporary storage measures and vocabulary recall. This discrepancy cannot be attributed to ceiling effects because accuracy in vocabulary recall was comparable across studies (Brooks & Kempe, 2013: 74%, current study: 76%). Given that tasks, constructions, and distributional characteristics of the input differed across studies, we cannot readily explain the differing patterns of results; nevertheless, the current findings suggest that memory capacity measures may not be robust predictors of vocabulary recall.

The role of nonverbal intelligence in accounting for individual differences in learning outcomes constitutes an additional challenge to the idea that implicit procedural learning underlies L2 learning of morphology. Whereas statistical-learning ability predicted accuracy in case production for old items, nonverbal intelligence predicted accuracy in case production for *all* items, with the effect mediated, for the most part, by explicit awareness (see also Brooks & Kempe [2013]). Thus, the picture that emerges from our results is a nuanced one that points towards co-existence of multiple learning mechanisms in adult L2-learners: both associative learning of item-based contingencies in the input and explicit category-based learning to support generalization. As

noted above in the discussion of the effect of the input manipulation, different mechanisms may subserve case production for old and new nouns, with item-based learning (i.e., associating suffixes with individual nouns) playing a key role in the correct retrieval of case marking for familiar nouns and category-based learning (i.e., associating sets of suffixes with gender categories) supporting generalization. Our results linked nonverbal intelligence with explicit noticing of the form-based regularities, i.e., with awareness that the forms of nouns in the nominative case provided cues to which suffixes to use in each semantic context.

In line with evidence from non-linguistic categorization tasks that individuals have different proclivities to engage in memorization versus rule abstraction (Little & McDaniel, 2015), language learners may differ in their predispositions to seek underlying regularities when learning complex morphological paradigms. Thus, the effect of nonverbal intelligence on case production might in part reflect individual differences in preference for rule learning. In this context, it is important to note that recent research calls into question the very idea that associative learning of contingencies can proceed entirely without explicit awareness (Weidemann et al., 2016), a suggestion that aligns with our observation that explicit awareness mediated the link between nonverbal intelligence and accuracy in learning and generalizing case marking.

Conclusions

This study used a miniature Russian case-marking paradigm to examine the impact of varying the structure of the input and individual differences in cognitive skills on L2 morphology learning. In support of the hypothesis that less predictable input supports generalization in language learning, we documented a short-lived benefit for balanced over skewed input, which was limited to the generalization of case-marking suffixes to new vocabulary in the first test (at the end of session 2). This transient effect of an input manipulation contrasted with the more

sustained role of individual differences in statistical-learning ability and nonverbal intelligence, which determined how well learners were able to memorize item-based contingencies between nouns and suffixes and extract form-based cues to the underlying gender categories—two distinct mechanisms implicated in learning and generalization of morphological patterns. These findings extend previous work on individual differences in L2 morphology learning by underscoring the critical role of pattern extraction and the robustness of individual differences in aptitude in the face of variable input.

Footnote

1. We ran parallel sets of analyses using proficiency in any L2 (an index of bilingualism) and proficiency in L2s with grammatical gender as predictors of language outcomes. Neither measure of proficiency was related to any outcome measure.

References

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012).

Determinants of success in native and non- native listening comprehension: An

individual differences approach. *Language Learning*, *62*(S2), 49-78.

DOI:10.1111/j.1467-9922.2012.00706.x

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language

learning device. *Psychological Review*, *105*(1), 158-173. DOI:10.1037/0033-

295X.105.1.158

Bates, E., & Goodman, J.C. (1997). On the inseparability of grammar and the lexicon: Evidence

from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*,

12(5/6), 507-584. DOI:10.1080/016909697386628

Brooks, P.J., Kempe, V. (2013). Individual differences in adult foreign language learning: The

mediating effect of meta-linguistic awareness. *Memory & Cognition, 41,* 281-296.

DOI:10.3758/s13421-012-0262-9

Brooks, P.J., Kempe, V., & Donachie, A. (2011). Second language learning benefits from

similarity in word endings: Evidence from Russian. *Language Learning*, *61*(4), 1142-

1172. DOI:10.1111/j.1467-9922.2011.00665.x

Brooks, P.J., Kempe, V. & Sionov, A. (2006). The role of learner and input variables in learning

inflectional morphology. *Applied Psycholinguistics*, *27,* 185-209.

DOI:10.1017/S0142716406060243

Casenhiser, D., & Goldberg, A.E. (2005). Fast mapping between a phrasal form and

meaning. *Developmental Science*, *8*(6), 500-508. DOI:10.1111/j.1467-7687.2005.00441.x

Cattell, R.B. & Cattell, H.E.P. (1973). *Measuring Intelligence with the Culture Fair Tests.* Champaign, IL: Institute for Personality and Ability Testing.

Cohen, J.D., MacWhinney, B., Flatt, M. & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments & Computers, 25*(2), 257-271. DOI:10.3758/BF03204507

Conway, A.R., Cowan, N., Bunting, M.F., Therriault, D.J., & Minkoff, S.R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163-183. DOI:10.1016/S0160-2896(01)00096-4

Conway, C.M., Bauernschmidt, A., Huang, S.S., & Pisoni, D.B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356-371. DOI:10.1016/j.cognition.2009.10.009

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450-466. DOI:10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422-433. DOI:10.3758/BF03214546

DeKeyser, R.M. (2005). What makes learning second- language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1-25. DOI:10.1111/j.0023-8333.2005.00294.x

Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. & Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*(5478), 457-460. DOI:10.1126/science.289.5478.457

Erickson, L.C., & Thiessen, E.D. (2015). Statistical learning of language: Theory, validity, and

   predictions of a statistical learning account of language acquisition. *Developmental

   Review, 37*(5), 66-108. DOI:10.1016/j.dr.2015.05.002

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*(2), 321-

   335. DOI:10.1044/1092-4388(2009/07-0189)

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy

   acquisition in a second language? *Psychological Science*, *24*(7), 1243-1252.

   DOI:10.1177/0956797612472207

Gebauer, G.F., & Mackintosh, N.J. (2007). Psychometric intelligence dissociates implicit and

   explicit learning. *Journal of Experimental Psychology: Learning, Memory, and

   Cognition*, *33*(1), 34-54. DOI:10.1037/0278-7393.33.1.34

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic

   categories. *Journal of Child Language*, *32*(2), 249-268.

   DOI:10.1017/S0305000904006786.

Goldberg, A.E., Casenhiser, D., & White, T.R. (2007). Constructions as categories of language.

   New Ideas in Psychology, *25*(2), 70-86. DOI:10.1016/j.newideapsych.2007.02.004

Goldberg, A.E., Casenhiser, D.M., & Sethuraman, N. (2004). Learning argument structure

   generalizations. *Cognitive Linguistics*, *15*(3), 289-316. DOI:0936–5907/04/0015–289

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5),

   431-436. DOI:10.1111/1467-9280.00476

Granena, G. (2013). Individual differences in sequence learning ability and second language

    acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665-703.

    DOI:10.1111/lang.12018

Grigorenko, E.L., Sternberg, R.J., & Ehrman, M.E. (2000). A theory-based approach to the

    measurement of foreign language learning ability: The Canal-F theory and test. *The*

    Modern Language Journal, *84*(3), 390-405. DOI: http://www.jstor.org/stable/330568

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and

    immediate serial recall in adults. *Quarterly Journal of Experimental Psychology (A):*

    Human Experimental Psychology, *56*(7), 1213-1236. DOI:10.1080/02724980343000071

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory:

    Computational and neural bases. *Brain and Language*, *59*(2), 267-333.

    DOI:10.1006/brln.1997.1819

Gwet, K.L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the*

    extent of agreement among raters. Gaithersburg, MD: Advanced Analytics.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in

    incidental language learning. *Learning and Individual Differences, 44*, 9-15.

    DOI:10.1016/j.lindif.2015.10.003

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and

    towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.

    DOI:10.1016/j.jml.2007.11.007

Kaufman, S.B., DeYoung, C.G., Gray, J.R., Jiménez, L., Brown, J., & Mackintosh, N. (2010).

    Implicit learning as an ability. *Cognition*, *116*(3), 321-340.

    DOI:10.1016/j.cognition.2010.05.011

Kempe, V., & Brooks, P.J. (2001). The role of diminutives in the acquisition of Russian gender: Can elements of child- directed speech aid in learning morphology? *Language Learning, 51*(2), 221-256.DOI: 10.1111/1467-9922.00154

Kempe, V. & Brooks, P.J. (2008). Second language learning of complex inflectional systems. *Language Learning, 54*(4), 703-746. DOI:10.1111/j.1467-9922.2008.00477.x

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171-184. DOI:10.1037/a0025405

Leow, R.P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26*(1), 35-57. DOI:10.1017/S0272263104261022

Leow, R.P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research, 30*(2), 111-127. DOI:10.1177/0267658313511979

Linck, J.A., Hughes, M.M., Campbell, S.G., Silbert, N.H., Tare, M., Jackson, S.R., ... & Doughty, C.J. (2013). Hi- LAB: A new measure of aptitude for high- level language proficiency. *Language Learning*, *63*(3), 530-566. DOI:10.1111/lang.12011

Linck, J.A., Osthus, P., Koeth, J.T. & Bunting, M.F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861-883. DOI:10.3758/s13423-013-0565-2

Little, J.L., & McDaniel, M.A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition,* 43, 283-297. DOI:10.3758/s13421-014-0475-1

Lum, J.A., Conti-Ramsden, G., Morgan, A.T., & Ullman, M.T. (2014). Procedural learning

    deficits in specific language impairment (SLI): A meta-analysis of serial reaction time

    task performance. *Cortex*, *51,* 1-10. DOI:10.1016/j.cortex.2013.10.011

Mandelbrot, B. (1953). An informational theory of the statistical structure of

    language. *Communication Theory*, *84*, 486-502.

Marchman, V.A., & Bates, E. (1994). Continuity in lexical and morphological development: A

    test of the critical mass hypothesis. *Journal of Child Language*, *21*(2), 339-366.

    DOI:10.1017/S0305000900009302

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. DOI:10.1111/j.1551-6709.2009.01091.x

McDonough, K., & Nekrasova-Becker, T. (2012). Comparing the effect of skewed and balanced

    input on English as a foreign language learners' comprehension of the double-object

    dative construction. *Applied Psycholinguistics*, *35*(2), 419-442.

    DOI:10.1017/S0142716412000446

McDonough, K., & Trofimovich, P. (2013). Learning a novel pattern through balanced and

    skewed input. *Bilingualism: Language and Cognition*, *16*(3), 654-662.

    DOI:10.1017/S1366728912000557

McLaughlin, B. (1980). Theory and research in second language learning: An emerging

    paradigm. *Language Learning*, *30*(2), 331-350. DOI:10.1111/j.1467-

    1770.1980.tb00322.x

Misyak, J.B., & Christiansen, M.H. (2012). Statistical learning and language: An individual

differences study. *Language Learning*, *62*(1), 302-331. DOI:10.1111/j.1467-

9922.2010.00626.x

Misyak, J.B., Christiansen, M.H., & Tomblin, J.B. (2010). On-line individual differences in

statistical learning predict language processing. *Frontiers in Language Sciences*,

1(00031). DOI:10.3389/fpsyg.2010.00031

Miyake, A., & Friedman, N.P. (1998). Individual differences in second language proficiency:

Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.),

Foreign language learning: Psycholinguistic studies on training and retention (pp. 339–

364). Mahwah, NJ: Erlbaum.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., & Wong, P.C.M.

(2014). Declarative and procedural memory as individual differences in second language

acquisition. *Bilingualism: Language and Cognition, 17*(1), 56-72.

DOI:10.1017/S1366728912000715

Nakamura, D. (2012). Input skewedness, consistency, and order of frequent verbs in frequency-

driven second language construction learning: A replication and extension of Casenhiser

and Goldberg (2005) to adult second language acquisition. *International Review of

Applied Linguistics in Language Teaching*, *50*(1), 1-37. DOI:10.1515/iral-2012-0001

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.*

Cambridge: Cambridge University Press.

Polio, C., & Shea, M.C. (2014). An investigation into current measures of linguistic accuracy in

second language writing research. *Journal of Second Language Writing*, *26*, 10-27.

DOI:10.1016/j.jslw.2014.09.003

Reber, A.S. (2015). Forward. In P. Rebuschat (Ed.) *Implicit and explicit learning of languages* (pp. vii-viii). Amsterdam: John Benjamins.

Reber, A.S., Walkenfeld, F.F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 888-896. DOI:10.1037/0278-7393.17.5.888

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368-392. DOI:10.1177/026765830101700405

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*(2), 235-268. DOI:10.1017/S0272263105050126

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. DOI:10.1126/science.274.5294.1926

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh: Psychology Software Tools.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, *45*(1), 21-50. DOI:10.1080/14640749208401314

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4-27. DOI:10.1037/0096-3445.125.1.4

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical

    perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105-120.

    DOI:10.1016/j.jml.2015.02.001

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name

    agreement, image agreement, familiarity and visual complexity. *Journal of Experimental*

    Psychology: Human Learning and Memory, 6(2), 174-215.

    DOI:10.1016/j.jml.2015.02.001

Speciale, G., Ellis, N.C., & Bywater, T. (2004). Phonological sequence learning and short-term

    store capacity determine second language vocabulary acquisition. *Applied*

    Psycholinguistics, *25*(2), 293-321. DOI:10.1017/S0142716404001146

Ullman, M.T. (2001). The neural basis of lexicon and grammar in first and second language: The

    declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105-122.

    DOI:10.1017/S1366728901000220

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural

    model. *Cognition*, *92*(1), 231-270. DOI:10.1016/j.cognition.2003.10.008

Weidemann, G., Satkunarajah, M., & Lovibond, P.F. (2016). I think, therefore eyeblink: The

    importance of contingency awareness in conditioning. *Psychological Science, 27*(4), 467-

    475. DOI:10.1177/0956797615625973

Year, J. & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive

    construction: The role of verb prototype, input distribution, and frequency. *Modern*

    Language Journal, *93*, 399-417. DOI:10.1111/j.1540-4781.2009.00898.x

Table 1. Russian question-and-answer dialogs for dative, genitive, and instrumental case. Literal

translations are provided in brackets.

	Question (same for both genders)	Gender	Examples of Answers
Dative	Kuda idjot slon? [To where is elephant going?]	feminine	k zvezde [towards star]
		masculine	k zaboru [towards fence]
Genitive	Otkuda ukhodit slon? [From where is elephant coming?	feminine	ot zvezdy [from star]
		masculine	ot zabora [from fence]
Instrumental	Gde pruachetsya slon? [Where is elephant hiding?]	feminine	pod zvezdoj [under star]
		masculine	pod zaborom [under fence]

Table 2: Descriptive statistics for predictor variables for the two input conditions and results of t-tests comparing input conditions.

| | | Input Condition | | | | | |
| | | Balanced | | Skewed | | | |
Predictor Variables	# Items on Test	Mean (SD)	Range	Mean (SD)	Range	$t(52)$	partial η^2
Total Number of Languages Learned		3.0 (.9)	2–5	3.1 (.9)	2–5	−.43	.004
Nonword Repetition	90	65.4 (12.5)	26–86	61.9 (8.0)	40–77	1.22	.028
Reading Span	70	45.6 (6.5)	31–55	44.3 (7.6)	29–61	.67	.009
Auditory AGL	40	23.3 (3.9)	15–33	24.2 (5.9)	17–40	−.66	.008
Culture-Fair Intelligence	50	23.5 (4.5)	15–31	24.1 (5.7)	15–35	−.43	.003

Table 3: Accuracy for outcome measures and results of t-tests comparing input conditions.

| | | Input Condition | | | | | |
| | | Balanced | | Skewed | | | |
		Mean (SD)	Range	Mean (SD)	Range	t(52)	partial η^2
Session 1							
Training	Case Comprehension	75.8% (20.7)	25.9–97.2	69.5% (26.9)	26.7–100	.97	.018
Session 2							
Training	Case Comprehension	91.6% (15.7)	29.0–100	86.8% (22.2)	37.4–100	.92	.016
Testing	Case Production (old items)	51.2% (20.7)	11.1–79.6	47.7% (23.1)	11.1–92.6	.62	.007
	Case Production (new items)	25.5% (15.8)	4.2–66.7	18.8% (16.6)	0–54.2	1.51	.042
	Vocabulary Recall	70.0% (13.8)	38.9–88.9	70.0% (21.6)	33.3–100	0	.000
Session 3							
Training	Case Comprehension	94.7% (12.7)	36.4–100	95.9% (9.6)	63.9–100	−.41	.003
Testing	Case Production (old items)	62.8% (21.5)	11.1–90.7	63.8% (21.3)	20.4–98.1	−.18	.001
	Case Production (new items)	36.0% (19.3)	0–75.0	34.6% (23.1)	0–95.8	.24	.001
	Vocabulary Recall	81.1% (14.0)	50.0–100	77.6% (19.3)	22.2–100	.76	.011
	Awareness	1.2 (.6)	0–2	1.1 (.8)	0–2	.38	.003

Table 4: Correlations between predictor variables.

	Nonword Repetition	Reading Span	Auditory AGL	Culture-Fair Intelligence
Total Number of Languages Learned	.03	−.12	.04	.08
Nonword Repetition		.30	.24	.23
Reading Span			.20	.15
Auditory AGL				.27

Table 5: Correlations between outcome variables.

	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Case Comprehension Session 1	.61*	.49*	.44*	.34	.33	.35	.35	.35	.40
2. Case Comprehension Session 2		.57*	.48*	.26	.78*	.52*	.43	.20	.21
3. Case Production (old items) Session 2			.85*	.43*	.50*	.85*	.73*	.30	.44*
4. Case Production (new items) Session 2				.45*	.36	.74*	.66*	.30	.42
5. Vocabulary Recall Session 2					.12	.40	.26	.76*	.18
6. Case Comprehension Session 3						.56*	.37	.03	.05
7. Case Production (old items) Session 3							.80*	.37	.33
8. Case Production (new items) Session 3								.33	.45*
9. Vocabulary Recall Session 3									.14
10. Awareness Session 3									

*significance after Bonferroni-correction at $p=.001$.

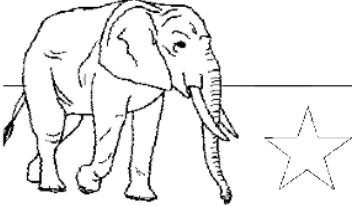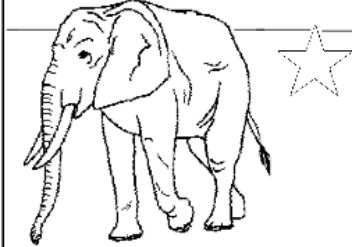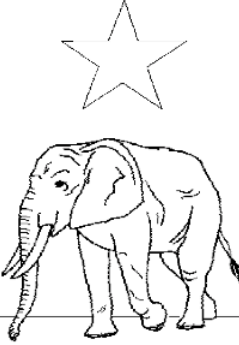Table 6: Factor loadings (>.30) obtained by Principal Component Analysis of the outcome measures.

	Factor 1 50.3% variance	Factor 2 15.1% variance	Factor 3 11.5% variance
Case Comprehension Session 1	.32	.47	−.36
Case Comprehension Session 2	--	.89	--
Case Production (old items) Session 2	.77	.46	--
Case Production (new items) Session 2	.76	.33	--
Vocabulary Recall Session 2	--	--	.90
Case Comprehension Session 3	--	.93	--
Case Production (old items) Session 3	.71	.48	--
Case Production (new items) Session 3	.81	--	--
Vocabulary Recall Session 3	--	--	.93
Awareness Session 3	.76	--	--

Table 7: Unstandardized coefficients and associated z-values with Input Condition (IC: balanced vs. skewed), Artificial Grammar Learning (AGL), and Culture-Fair Intelligence (CFI) as predictors in a mixed-effects multiple regression model. Log odds ratios are provided in parentheses to indicate the change in the odds per unit of each predictor when controlling for the other predictors.

.

		Session 1	Session 2				Session 3			
		Case Comp1	Case Comp2	Case Production2 (old items)	Case Production2 (new items)	Vocabulary Recall2	Case Comp3	Case Production3 (old items)	Case Production3 (new items)	Vocabulary Recall3
IC	B	.17	−.01	.12	.35	.03	−.35	−.02	.09	.12
		(1.19)	(0.99)	(1.13)	(1.42)	(1.03)	(0.70)	(0.98)	(1.09)	(1.13)
	z	.88	−.03	1.02	2.37^{*}	.22	−1.08	−.12	.55	.71
AGL	B	.57	.78	.38	.20	−.11	.68	.29	.18	−.02
		(1.77)	(2.18)	(1.46)	(1.22)	(0.90)	(1.97)	(1.34)	(1.20)	(0.98)
	z	2.68^{**}	2.13^{*}	3.10^{**}	1.31	−.70	$1.76^{\dagger}$	$1.93^{\dagger}$	1.08	−.11
CFI	B	.57	.64	.34	.48	−.19	.57	.32	.44	−.09
		(1.77)	(1.90)	(1.40)	(1.62)	(0.83)	(1.77)	(1.38)	(1.55)	(0.94)
	z	2.80^{**}	$1.88^{\dagger}$	2.78^{**}	3.07^{**}	−1.28	$1.73^{\dagger}$	2.20^{*}	2.61^{**}	−.53
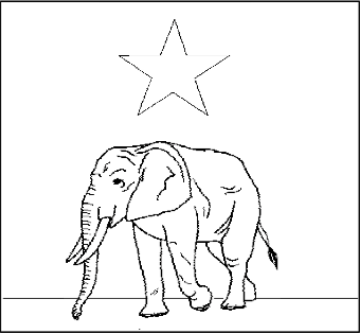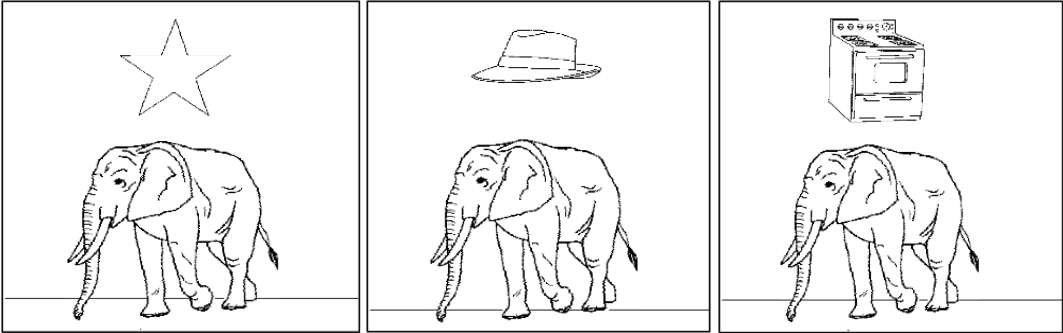
$^{***}p<.001, ^{**}p<.01, ^{*}p<.05, ^{\dagger}p<.1$

Figure 1. Example set of items for nominative, dative, genitive, and instrumental case.
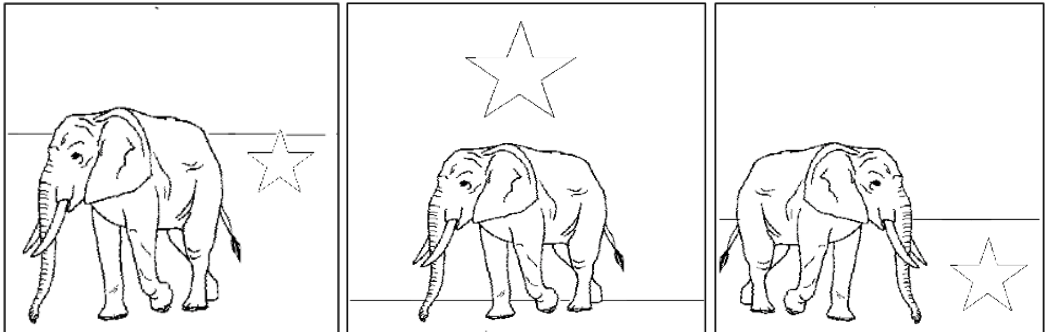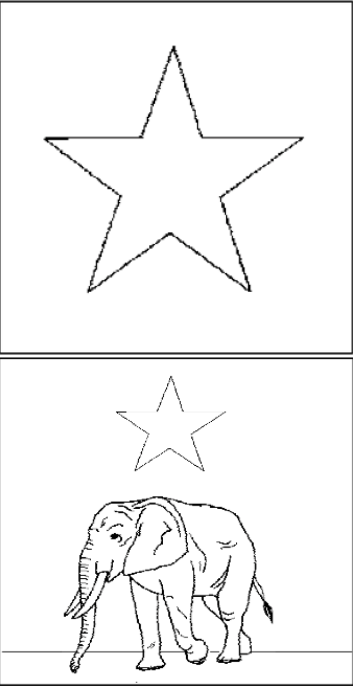
eto zvezda	*k zvezde*	*ot zvezdy*	*pod zvezdoj*
[this is star(fem)–NOM]	[towards star(fem)–DAT]	[from star(fem)–GEN]	[under star(fem)–INS]

Note: fem=feminine, NOM=nominative, DAT=dative, GEN=genitive, INS=instrumental

Figure 2. Examples of trials for each training task.

Task and Dialog	Picture(s) presented	Instructions
Listen and Repeat Male voice: *Gde pruachetsya slon?* [Where is elephant hiding?] Female voice: *pod zvezdoj* [under star]		Listen to the dialog and repeat the woman's answer to the question. (Answer is replayed) Participant is instructed to repeat the Russian phrase a second time to advance to the next trial.
Noun Comprehension Male voice: *Gde pruachetsya slon?* [Where is elephant hiding?] Female voice: *pod zvezdoj* [under star]		Listen to the dialog and select the picture that matches what the woman says. (After response) Russian phrase (*pod zvezdoj*) is replayed as correct picture is shown. Participant is instructed to repeat the Russian phrase to advance.

Case Comprehension Male voice: *Gde pruachetsya slon?* [Where is elephant hiding?] Female voice: *pod zvezdoj* [under star]		Listen to the dialog and select the picture that matches what the woman says. (After response) Russian phrase (*pod zvezdoj*) is replayed as correct picture is shown. Participant is instructed to repeat the Russian phrase to advance.
Case Production Female voice: *Eto zvezda* [This is star.] Male voice: *Gde pruachetsya slon?* [Where is elephant hiding?]		Listen to the question and answer it in Russian. (After response, correct phrase is played.) Female voice: *pod zvezdoj* [under star] Participant is instructed to repeat the correct phrase to advance.

Appendix: Stimuli

	Transliteration	Gloss	Set	Frequency	
				Balanced	Skewed
Feminine nouns					
буква	bukva	letter	Training	6	18
юла	yula	top	Training	6	12
плита	plita	stove	Training	6	6
шуба	shuba	fur-coat	Training	6	3
шляпа	shlyapa	hat	Training	6	3
штора	shtora	curtain	Training	6	3
звезда	zvezda	star	Training	6	3
луна	luna	moon	Training	6	3
метла	metla	broom	Training	6	3
тыква	tykva	pumpkin	Test 1		
кофта	kofta	blouse	Test 1		
труба	truba	chimney	Test 1		
игла	igla	needle	Test 1		
морда	morda	snout	Test 2		
парта	parta	school-desk	Test 2		
пила	pila	saw	Test 2		
гора	gora	mountain	Test 2		
Masculine nouns					
поезд	poyezd	train	Training	6	18
забор	zabor	fence	Training	6	12
бокал	bokal	goblet	Training	6	6
мусор	musor	garbage	Training	6	3
остров	ostrov	island	Training	6	3
город	gorod	city	Training	6	3
диван	divan	couch	Training	6	3
кувшин	kuvshin	jug	Training	6	3
тюльпан	tyulpan	tulip	Training	6	3
пояс	poyas	belt	Test 1		
череп	cherep	skull	Test 1		
халат	khalat	bathrobe	Test 1		
арбуз	arbuz	watermelon	Test 1		
галстук	galstuk	tie	Test 2		
провод	provod	wire	Test 2		
конверт	konvert	envelope	Test 2		
стакан	stakan	glass	Test 2		