Applied Psycholinguistics **27** (2006), 185–209
Printed in the United States of America
DOI: 10.1017.S0142716406060243

The role of learner and input variables in learning inflectional morphology

PATRICIA J. BROOKS
College of Staten Island and Graduate Center, City University of New York

VERA KEMPE
University of Stirling

ARIEL SIONOV
College of Staten Island, City University of New York

Received: April 19, 2005 Accepted for Publication: July 18, 2005

ADDRESS FOR CORRESPONDENCE
Patricia Brooks, Department of Psychology, 4S-108, College of Staten Island, City University of New York, 2800 Victory Boulevard, Staten Island, NY 10314. E-mail: pbrooks@mail.csi.cuny.edu

ABSTRACT
To examine effects of input and learner characteristics on morphology acquisition, 60 adult English speakers learned to inflect masculine and feminine Russian nouns in nominative, dative, and genitive cases. By varying training vocabulary size (i.e., type variability), holding constant the number of learning trials, we tested whether learners required a "critical mass" of vocabulary to generalize case marking patterns to new nouns. Cattell's Culture-Fair IQ Test mediated the effect of type variability on success in generalizing case marking to new vocabulary: only participants with above-median Culture-Fair Test scores showed the predicted critical mass effect of better generalization with larger training vocabulary. These results demonstrate how individual differences in central executive functioning and attention allocation capacity can affect adult second language learning.

Every natural language encompasses a highly complex system of categories at multiple levels of linguistic organization (Gomez & Gerkin, 2000; Lakoff, 1987). Language learners are faced with the task of organizing linguistic input in terms of phonological, semantic, pragmatic, lexical, morphological, and syntactic distinctions. This categorization problem is especially challenging to adult second language (L2) learners, who may be acquiring L2 categories based on limited input, and with potential interference from their knowledge of other languages (Frenck-Mestre, Foucart, & Caetano-Nunes, 2004; Gesi Blanchard, 1998; Hernandez, Li, & MacWhinney, 2005; MacWhinney, 1992). With respect to morphological categorization, languages vary enormously in the richness of their inflectional patterns. English, in particular, presents a very impoverished set of noun and verb inflections, relative to many other languages (e.g., Russian, Hungarian, Lithuanian, Finnish, Spanish). One feature of natural languages that is especially difficult for adult learners involves mastery of grammatical dependencies (Braine,

© 2006 Cambridge University Press 0142-7164/06 $12.00

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

1987; Braine et al., 1990; Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Newport & Aslin, 2004), such as agreement patterns (e.g., subject–verb agreement, adjective–noun agreement), noun declensions, and verb conjugations.

In this paper, we examine several of the factors responsible for success in learning complex morphological structures, using Russian noun declension as a representative domain for exploration. The literature on first language (L1) and L2 acquisition has proposed various characteristics of both the input and the learner to be predictive of success in acquiring grammatical dependencies. On the input side, variability in vocabulary (i.e., greater type frequency) has been proposed to facilitate attention to grammatical regularities, and to provide a broader database for generalizing grammatical patterns to new vocabulary (Bates & Goodman, 1997; Bybee, 1995; Bybee & Hopper, 2001; Ellis, 2002; Gomez, 2002; Tomasello, 2003). On the side of the learner, cognitive factors such as phonological loop capacity, verbal working memory (VWM), and the ability to focus attention to cope with novelty and ambiguity have been invoked to explain individual differences in learners' success (e.g., Ellis & Schmidt, 1998; Grigorenko, Sternberg & Ehrman, 2000; Miyake & Friedman, 1998; Winke, 2005). Below, we briefly present the proposed contributions of input and learner-based variables to success in learning complex morphosyntactic structures.

THE ROLE OF INPUT

Elizabeth Bates and many other researchers have emphasized the close relationship between lexical and grammatical development (e.g., Bates & Goodman, 1997; Dionne, Dale, Boivin, & Plomin, 2003; Plunkett & Juola, 1999). One interpretation of this relationship is that acquiring a critical mass of vocabulary is crucial to learning the grammar of a language, because input variability provides the learner with a sufficient database from which to extract morphosyntactic regularities. In the domain of morphology, Marchman and Bates (1994) showed that the learning of English past tense forms was tightly correlated with the size and distribution of children's verb vocabularies at all points in development. Gomez (2002) provided experimental evidence for this relationship by using a miniature artificial language to test how variability in vocabulary contributes to learning nonadjacent dependencies. Learners were required to note patterns of co-occurrence between monosyllabic "grammatical" elements separated by intervening bisyllabic words (e.g., that *pel* co-occurred with *rud*, and *vot* co-occurred with *jic* in strings like *pel wadim rud* or *vot wadim jic*). Gomez observed, in both infant and adult participants, that increasing vocabulary size, and thereby decreasing predictability between grammatical elements and adjacent words, led to better learning of the nonadjacent dependencies. Even though the materials in the Gomez study consisted of artificially constructed nonwords, the task is relevant with respect to learning of inflectional morphology. For example, in many languages, the learner has to acquire the association between a preposition before the noun and a case inflection at the end of the noun such as in the Russian *ot zabora* [from fence-genitive] or *k zaboru* [to fence-dative]. Her finding provides crucial empirical evidence for the view that learning morphosyntactic regularities requires a critical mass of vocabulary, and that detecting nonadjacent dependencies is contingent

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Table 1. *Summary of dialogues used in language learning task*

Case	Gender	Question (Same for Both Genders)	Answer (Example)
Nominative	Masc.	Cto eto? (What is this?)	Eto zabor. (This is fence.)
	Fem.		Eto zvezda. (This is star.)
Dative	Masc.	Kuda idjot slon? (To where is	K zaboru. (To fence.)
	Fem.	elephant going?)	K zvezde. (To star.)
Genitive	Masc.	Otkuda ukhodit slon? (From where is	Ot zabora. (From fence.)
	Fem.	elephant going?)	Ot zvezdy. (From star.)

upon a sufficient amount of variability of the segments located between the critical elements.

The present study tests whether similar benefits of a larger training vocabulary on the acquisition of grammatical regularities can be observed in learners exposed to the case marking patterns of a natural language (i.e., Russian), which are considerably more complex than the grammatical dependencies studied by Gomez (2002). Learners of Russian not only must acquire nonadjacent dependencies between prepositions and case markers, but also must learn noun subclasses (i.e., declension paradigms), which take different sets of case markers. For example, when encountering a noun ending in *-a*, for example, *zvezda* [star], and the question "to where the elephant is going?" the learner needs to select the preposition *k*, and one of two dative suffixes, namely the feminine dative suffix *-e*, to correctly produce *k zvezde* [to star]. (See Table 1 for a summary of the portion of the Russian declension system used in the present experiment.) This learning is further complicated by the existence of allomorphy (i.e., the exact forms of prepositions and case markers are influenced somewhat by the phonological shapes of adjacent words). Thus, an important goal of the present study is to determine whether the same beneficial effect of type variability on grammar learning, as observed by Gomez (2002), will obtain when the language to be learned is not an artificial one, but is part of a real human language.

LEARNER VARIABLES

In addition to variables related to the input, successful learning of morphosyntactically complex patterns is dependent on the cognitive abilities of the learner. Adult L2 learners differ considerably in their success in extracting and generalizing inflectional patterns from the input to form linguistic categories (R. Ellis, 1994; Skehan, 1989). Individual differences in language-learning abilities are more prominent in research on L2 learning, in comparison to studies of L1 acquisition. Recently, a number of researchers have examined VWM as a predictor of individual differences in acquiring L2 morphology and syntax. Thus, to learn the grammatical dependencies discussed above, learners must be able to maintain representations of linguistic forms over successive syllables or words, and VWM may play a crucial role in this process. Ellis and Schmidt (1998), using an artificial

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

language-learning paradigm, demonstrated that both short-term and long-term memorization ability predicted grammaticality judgments for sentences with distant dependencies (i.e., subject–verb agreement for transitive sentences with subject–object–verb word order), but neither predicted learning of more local dependencies (adjective–noun agreement and subject–intransitive verb agreement). Miyake and Friedman (1998) found links between VWM, measured through performance on the Reading Span task (Daneman & Carpenter, 1980), and syntactic comprehension ability in Japanese learners of English. Specifically, they argued that learning long-distance cues to grammatical roles, such as word order and subject–verb agreement, relies more heavily on VWM than learning local cues. Thus, the literature on L2 consistently points toward the role of VWM in learning long-distance associations between morphosyntactic elements.

The literature on the relationship between VWM and language learning describes several distinct measures of VWM. Studies on adult language learners traditionally use the Reading Span task introduced by Daneman and Carpenter (1980). Studies concerned with the effects of VWM on L1 acquisition in children have used nonword repetition and nonword memory span tasks as measures of VWM capacity (e.g., Adams & Gathercole, 2000; Gathercole, Hitch, Service, & Martin, 1997). Although each of these tasks is assumed to constitute a measure of VWM capacity, they differ with respect to the underlying components examined. The nonword repetition and memory span tasks are generally interpreted as direct measures of phonological loop capacity (Baddeley, Gathercole, & Papagno, 1998), whereas the Reading Span task has been shown to involve executive as well as storage components (Kane & Engle, 2002). Increasingly, working memory is viewed as being crucial in the regulation of cognitive activities, which fundamentally involves the control or focusing of attention (Cowan, 2000). With respect to L2 acquisition, Robinson (1995) has suggested that the more complex the material to be learned, the more one's attentional capacity will be predictive of language learning outcome.

We suspect that a sizeable proportion of the variability in individual differences in L2 learning may involve the executive component of VWM. More specifically, the allocation of attentional resources is crucial for organizing and managing concurrent task requirements in complex tasks (Duncan, Emslie, Williams, Johnson, & Freer, 1996). This is of direct relevance for language learning: if learners are exposed to natural language input, learning always involves multiple aspects like retaining novel vocabulary while at the same time registering a wide array of grammatical regularities. On this basis, we would predict that executive functioning would be related to successful L2 learning in addition to pure phonological storage capacity. However, direct measures of executive functioning have never been explored in the context of morphology learning, thus making it impossible to determine the independent contribution of executive functioning in this task. To fill this gap, the present study uses Cattell's Culture-Fair Nonverbal Intelligence Test (Cattell, 1971) to obtain a more direct measure of executive functioning, in addition to using measures of nonword retention and Reading Span as indices of phonological memory and VWM capacity, respectively. The Culture-Fair Test is considered to be a measure of fluid intelligence associated with Spearman's g (Cattell & Cattell, 1973) that has been shown to correlate with executive

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

functioning (Duncan et al., 1996) as well as with language-learning aptitude (Grigorenko et al., 2000).

Although processing speed has been shown to be a major determinant of working memory capacity (Fry & Hale, 1996, 2000), working memory, in turn, tends to be correlated with fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Fry & Hale, 1996, 2000; Just & Carpenter, 1992). Just and Carpenter (1992) have suggested that the more representations, inferences, expectations, and problem-solving goals an individual is able to actively maintain, the better he or she is able to detect and compare patterns and relationships. Thus, the availability of activation resources explains much of the shared variance between working memory tests and tests of fluid intelligence. In addition, the efficient allocation of these resources by the central executive plays a role both in working memory tests and tests of fluid intelligence (Engle et al., 1999). However, whereas working memory tests and tests of fluid intelligence both tap into executive functioning to a certain degree, it has been suggested that executive functioning explains most of the variance in fluid intelligence that is not explained by individual differences in working memory capacity (Dempster, 1992).

At present, the contributions of working memory and fluid intelligence to individual differences in language learning are only beginning to be fully explored, and thus far, very little research has considered these factors in relation to the acquisition of complex inflectional systems, which are characteristic of many natural languages. The present study explores the contribution of the learner variables of phonological memory, VWM capacity, and executive functioning, in addition to the input variable of type variation in vocabulary, to the learning of complex morphological structures. The interesting question is whether individual differences can explain to what extent learners are able to make use of type variation in the input. We use a subset of the Russian declension system to examine the role of learner and input variables in learning inflectional morphology, testing the learning of Russian case marking by adult native speakers of English.

THE RUSSIAN CASE-MARKING SYSTEM

In what follows, we describe features of the complex Russian inflectional system that are relevant for the present study. Russian nouns are marked for gender, number, and case. Russian has three genders, masculine, feminine, and neuter, which can often be predicted by the morphophonological shape of the noun in the nominative case: masculine nouns are marked by a zero morpheme and therefore end in consonants. Most neuter nouns end in the suffix *-o* or its allomorphs. The majority of feminine nouns end in *-a* or its allomorph. There is also considerable complexity in case and number marking. Russian has six cases: nominative, accusative, genitive, dative, instrumental, and locative, which are marked differently for singular and plural in three declension paradigms. The exact form of the case marking depends on the gender of the noun. "Transparently" gender-marked feminine nouns (i.e., those that end in *-a* or its allomorph in the nominative case) follow the first declension, most masculine and neuter nouns follow the second

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

declension paradigm, and "nontransparently" gender-marked feminine nouns (i.e., those that end in a palatalized consonant in the nominative case) follow the third declension. In addition, Russian case marking varies with respect to animacy, which affects the marking of the accusative case in masculine nouns. Finally, there is a considerable degree of allomorphy such that the phonological shape of the noun determines the specific variant of the preposition and case marker.

The present experiment examines the acquisition of a portion of this inflectional system, focusing on masculine and feminine nouns in the nominative, genitive, and dative cases. This simplification is necessary for adults to learn any of the basic features of Russian noun morphology in a limited number of language learning sessions. The learners were presented with a picture of the semantic relationship depicted by an elephant moving toward or from an object, and had to answer the question as to where the elephant was going to or where the elephant was coming from. The direction of the movement determined the choice of preposition *ot* [from] (for genitive case) or *k* [to] (for dative case). The specific genitive or dative suffix depends on the gender of the noun: masculine nouns take *-a* in the genitive and *-u* in the dative, feminine nouns take *-y* or its allomorph *-i* in the genitive and *-e* in the dative. Thus, the task that the learners face is a complex one: they have to associate the semantic relationship with the correct preposition, learn which case is required for this preposition, and determine the gender of the noun to select the appropriate suffix.

METHOD

Participants

Sixty adults (43 women, 17 men, age range = 17–44 years) were recruited from Introductory Psychology classes at the College of Staten Island, CUNY, and received extra credit for their participation. None of the participants had any prior experience learning Russian, or any other Slavic language. Participants were pseudorandomly assigned to the three training vocabulary size conditions: because our participants varied considerably in terms of their prior experience with foreign languages, we administered a foreign language background questionnaire and matched participants in the vocabulary size conditions with respect to this measure, resulting in 20 participants with roughly equal numbers of men per condition.

Materials

Eighteen masculine and 18 feminine Russian nouns served as stimuli. These items were used to assemble training vocabularies of 6, 12, or 24 nouns, as well as a testing set of 12 nouns for assessing generalization. The three training sets and the testing set are presented in Appendix A. Each noun was associated with a corresponding line drawing, most of which were taken from the Snodgrass and Vanderwart (1980) set of standardized pictures. Line drawings of each object were presented separately (i.e., for nominative case) and in conjunction with a picture

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

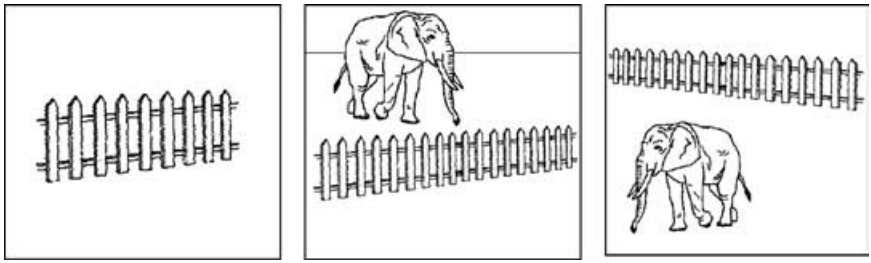


Figure 1. An example of a set of pictures used in the language learning task for nominative, dative, and genitive cases.

of an elephant walking toward the object (i.e., for dative case) or away from the object (i.e., for genitive case). Figure 1 presents an example set of pictures.

Procedure

Language learning task. Participants were tested individually and completed six sessions of training, followed by the test phase. The six training sessions were administered within a span of 14 days, and each lasted 45–60 min. The testing followed the last training session after a short break, and lasted about 15 min. The entire experiment was controlled using PsyScope experimental software (Cohen, MacWhinney, Flatt, & Provost, 1993).

During training and testing, all Russian phrases were presented aurally; participants never saw any words or phrases written in Russian. Participants in each vocabulary size condition received the same numbers of trials per block. What differed across training conditions was the number of repetitions of items within each block: participants in the size 24 condition heard one repetition, participants in the size 12 condition heard two repetitions, and participants in the size 6 condition heard four repetitions of each item per block. Order of trials was randomized for each condition.

Each training session comprised four blocks utilizing different tasks. The tasks were designed to engage learners in different activities while receiving exposure to phrases exemplifying the Russian case marking paradigm.

In Block 1 (Listen and Repeat Task), participants viewed pictures while listening to short dialogues that described the pictures. Each dialogue consisted of a question posed by a male speaker, followed by an answer spoken by a female speaker. Participants were instructed to repeat the answer as spoken by the female, which contained a noun (masculine or feminine) inflected for case (nominative, dative, or genitive). After the participant repeated this phrase, the correct pronunciation was presented again and the participant was asked to repeat it a second time. Participants' responses were tape-recorded. Block 1 comprised 72 randomized trials (24 nominative, 24 dative, 24 genitive). Table 1 presents examples of questions and answers for masculine and feminine nouns for each of the three cases.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

In Block 2 (Noun Comprehension Task), participants heard the same dialogues as in Block 1 and were tested on their comprehension of the nouns used in the answers. On each trial, participants viewed two pictures side by side. Both pictures depicted events corresponding to the same case (e.g., for dative case, both pictures showed an elephant walking toward an object). The two pictures differed in terms of which objects were shown. After hearing the dialogue, the participant was instructed to select the picture corresponding to the object mentioned in the dialogue, by pressing the corresponding button on a button box (i.e., left button for left picture, right button for right picture). After the participant made a choice, the phrase containing the case-inflected noun was presented again at the same time as the correct picture was shown. Block 2 comprised 72 randomized trials (24 nominative, 24 dative, 24 genitive). On each trial, the two pictures depicted training vocabulary items of the same gender.

In Block 3 (Case Comprehension Task), participants heard the dialogues for dative and genitive cases and were tested on their comprehension of the Russian prepositions and case marking. Two pictures were shown side by side. Both pictures depicted events involving the same object (e.g., a fence). The picture on the left depicted the elephant walking away from the object (genitive case), and the picture on the right depicted the elephant walking toward the object (dative case). The participant was instructed to listen carefully to the dialogue to find out whether the elephant moves away from or toward the object. After participants made a choice by pressing the corresponding button on a button box (i.e., left button for left picture, right button for right picture), the phrase containing the preposition and case-inflected noun was presented again at the same time as the correct picture was shown. Block 3 comprised 48 randomized trials (24 dative, 24 genitive).

In Block 4 (Case Production), participants' mastery of dative and genitive case marking patterns was probed by requiring them to produce the answers to the previous dialogues. First, to remind the participant of the noun to be case inflected, the picture of the object was presented along with a phrase spoken by the female introducing the noun in the nominative case (e.g., *Eto zabor* [This is fence]). Second, a picture of the elephant going away from or toward the object was shown, and the participant heard the male voice ask the question "To where/From where is elephant going?" After the participant produced a response, the correct response was presented alongside the corresponding picture, and the participant was instructed to repeat the correct answer. Participants' responses were tape-recorded. Block 4 comprised 48 randomized trials (24 dative, 24 genitive).

The procedure for the testing phase was identical to Block 4 of training (case production) except that the 12 new nouns (6 masculine, 6 feminine) were added to test participants' ability to transfer case marking patterns to unfamiliar nouns. Each new noun was tested in both dative and genitive cases, so that the final test comprised 72 trials (48 old items, 24 new items) presented in a randomized order.

At the end of the testing phase, we administered a vocabulary test. The pictures of objects (i.e., the pictures corresponding to nominative case) were presented one at a time (6, 12, or 24 items, depending on the vocabulary size condition) and

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

the participant was asked to name each picture. No feedback was given. If the participant could not retrieve the name of the picture, they were instructed to say "I don't know" to proceed to the next item.

Measures of individual differences. We administered three tests of individual differences in VWM, phonological memory, and nonverbal intelligence. These tests were administered after the language learning task in sessions 2–4. Order of tests was randomized. VWM was assessed with the Reading Span Task, using materials and procedure as outlined in Daneman and Carpenter (1980). Participants were asked to read aloud a set of unrelated sentences, with each sentence printed individually on an index card. Sentences were read one at a time. After reading a set of sentences, participants were instructed to recall the final word of each sentence in the set. Sentences were presented in sets of increasing size starting with a size of two sentences per set up to a size of five sentences per set. There were 70 sentences in the test, comprising five sets at each set size (two to five sentences). There are two methods of scoring that have been used for the Reading Span task. The first method determines the span (two to five) by checking if the last words have been reproduced correctly for all sentences on a given span level. If participants complete only three out of five items correctly, 0.5 span level is added. The second method simply involves counting the number of correctly reproduced words out of the total of 70. The latter measure has been shown to correlate highly with the former (Shah & Miyake, 1996), but suffers less from a restriction of range, and was therefore chosen for the present analyses.

Phonological memory was assessed using a Non-word Span Task that followed procedures outlined in Baddeley et al. (1998). Participants heard sequences of Russian monosyllabic words repeated in either the same or transposed order. (None of these Russian words served as stimuli in the language learning task.) The sequences of Russian words increased in length over successive trials, and the participants were asked to identify the two sequences as either the same (e.g., *vid*, *trud, bog, sneg . . . vid, trud, bog, sneg*) or as different (e.g., *luch, cvet, boj, khleb . . . luch, boj, cvet, khleb*). In the different trials, the order of two adjacent Russian words was reversed. Sequences increased in length starting with a size of three words per sequence up to a size of eight words per sequence. Eight randomized trials were presented at each length (four in the same order, four in a different order) for a total of 48 trials.

Participants also completed the Cattell Culture-Fair Test of Nonverbal Intelligence, Scale 3, Form A (Cattell, 1971). Students were given four sets of multiple-choice problems. Each test set started with several example problems, and then progressed with problems of increasing difficulty. Participants were instructed to solve as many problems as they could in the allotted time (ranging from 2.5 to 4 min/problem set). Two of problem sets (series and matrices) involved selecting an abstract geometric stimulus (from six alternatives) to complete a series or pattern (matrix). One problem set (classification) required the participant to identify which two out of five stimuli were alike in some way (i.e., different from the other three). The last problem set (conditions/topology) required the participant to select a stimulus (out of five alternatives) that matched a standard condition with respect to the placement of a dot among geometric forms.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Table 2. *Percentage of correct case inflections for old and new items as a function of vocabulary size group (N = 20/group)*

	Vocabulary Size		
	6	12	24
Old items	88.5 (10.2)	78.4 (14.9)	66.5 (25.7)
New items	42.1 (16.3)	40.8 (18.3)	47.3 (27.9)

Note: Standard deviations are in parentheses.

RESULTS

Overall task performance

The crucial outcome measure in this study is participants' performance in the testing session administered after the six training sessions. Responses were analyzed with respect to whether participants produced the correctly inflected form of the noun.[1] Responses were coded as errors if participants used the wrong noun, used the wrong preposition (e.g., *ot zabora* instead of *k zaboru*) or repeated the phrase containing the noun in nominative case (e.g., *eto zabor* instead of *k zaboru*), or used an incorrect suffix (e.g., *k zabore* instead of *k zaboru*) or no suffix at all as in the masculine nominative that ends in a consonant (e.g., *k zabor* instead of *k zaboru*).

Table 2 summarizes case-marking accuracy for old and new items in the testing session for each of the three vocabulary size conditions. A 2 (Familiarity: old vs. new items) × 3 (vocabulary size: 6, 12, 24) mixed-type analysis of variance (ANOVA) conducted on the percentage of trials correct yielded a main effect of familiarity, $F\,(1, 57) = 293.2$, $p < .001$, as well as a significant interaction between familiarity and vocabulary size, $F\,(2, 57) = 16.0$, $p < .001$. This indicates that, at the end of training, accuracy in producing correct case marking for the old items was much greater than for the new items. To tease apart the interaction, separate one-way ANOVAs were conducted for old and new items. For old items, there was a significant effect of vocabulary size condition, $F\,(2, 57) = 7.4$, $p < .01$. Performance in the Size 24 condition was significantly worse than in the Size 6 and the Size 12 conditions (Fisher's protected least significant difference [LSD]: Size 6 vs. Size 24, $p < .001$; Size 12 vs. Size 24, $p < .05$). This effect of vocabulary size condition for old items is not altogether surprising, and merely indicates that participants in the Size 24 condition, who had more items to master, performed worse than those exposed to smaller training sets with more item repetition. Clearly, having to remember the inflections for 24 items, given many fewer repetitions, was difficult for our participants!

If a critical mass of vocabulary is crucial for enabling learners to generalize case marking patterns to new items then we would expect an effect of vocabulary size in the new items such that larger input vocabularies are associated with better performance. This is not what the data show. A one-way ANOVA testing the effect of condition for just the new items revealed no significant

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Table 3. *Means, standard deviations, and ranges of the three measures of individual differences for each vocabulary size group (N = 20/group)*

	M	*SD*	Range
Vocabulary size = 6			
Non-word Span (out of 40)	29.2	2.9	25–35
Reading Span (out of 70)	46.1	12.9	20–67
Cattell Culture-Fair Test	23.6	4.6	15–31
Vocabulary size = 12			
Non-word Span (out of 40)	30.1	3.4	20–34
Reading Span (out of 70)	46.8	5.6	39–63
Cattell Culture-Fair Test	22.6	4.9	14–31
Vocabulary size = 24			
Non-word Span (out of 40)	29.3	3.6	19–35
Reading Span (out of 70)	43.8	10.8	21–66
Cattell Culture-Fair Test	23.6	6.0	9–33

effect ($F < 1$). This finding is at odds with the results of Gomez (2002), who observed a clear advantage of larger vocabulary size on learning grammatical dependencies in an artificial language. In the next section, we will explore whether this null effect arises from collapsing over learners with very different abilities.

Exploring individual differences

Table 3 presents the means, standard deviations, and ranges for the measures of Non-word Span, Reading Span, and Culture-Fair IQ in the three vocabulary size groups. One-way ANOVAs with vocabulary size as between-subjects variable showed that the three groups did not differ with respect to any of the individual differences measures (all $Fs < 1$). We then computed correlations between the three measures for the entire sample, which revealed only a significant correlation between Non-word Span and Culture-Fair IQ Test scores ($r = .33$, $p < .05$, $N = 60$), suggesting that both measures tap into some shared component. We will return to this somewhat unexpected finding below. The correlation between Reading Span and Culture-Fair IQ test scores failed to reach significance ($r = .13$, $N = 60$, $p = .31$), as did the correlation between Reading Span and Non-word Span ($r = .19$, $N = 60$, $p = .16$).

To gain a better understanding of the participants' overall difficulty in generalizing case agreement to new items, we explored the relationship between the three measures of individual differences (Non-word Span, Reading Span, and Culture-Fair IQ), and performance in the testing session. Table 4 presents the correlations between the individual differences measures and performance on old and new items for each of the three vocabulary size conditions. These correlations show that both Reading Span and Culture-Fair IQ are positively correlated with performance on the old and new items in the Size 24 condition. This result suggests that in the Size 6 and 12 conditions the amount of noun

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Table 4. *Correlations between performance on old and new items and the measures of individual differences in the three vocabulary size groups (N = 20/group)*

	Vocabulary Size		
	6	12	24
Old items			
Non-word Span	.28	.00	.26
Reading Span	.15	−.12	.69**
Culture-Fair IQ	.30	.17	.67**
New items			
Non-word Span	−.05	.00	.38
Reading Span	−.16	−.19	.53*
Culture-Fair IQ	.24	.29	.69**

*$p < .05$. **$p < .01$.

exemplars provided was not sufficient to learn and generalize the case marking regularities, regardless of IQ or VWM capacity. However, performance in the Size 24 condition suggests that a larger vocabulary alone is not sufficient to generalize the regularities to new items and that certain individual prerequisites, as measured by Culture-Fair IQ and Reading Span, are required to be able to make use of the input.

To investigate what these individual prerequisites might be, and to determine the unique contributions of each variable over and above the contributions of the other variables, we performed a simultaneous regression with Reading Span and Culture-Fair IQ as predictor variables in the Size 24 condition. This analysis showed that both variables made independent contributions to performance in the old items (Reading Span: $\beta = .50$, $p < .01$; Culture-Fair IQ: $\beta = .47$, $p < .01$), but that Reading Span did not make an independent contribution over and above Culture-Fair IQ in the new items (Reading Span: $\beta = .30$, $p = .10$; Culture-Fair IQ: $\beta = .57$, $p < .01$). Because Reading Span and Culture-Fair IQ share components associated with attention allocation and executive functioning, these findings can be interpreted as follows: the independent contributions of Reading Span and Culture-Fair IQ in predicting performance on the old items in the Size 24 condition suggest that both the executive component measured by the Culture-Fair IQ Test as well as the storage component measured by the Reading Span task contribute to task performance. Thus, learning the case marking paradigm of 24 nouns over six sessions requires both storage capacity as well as the ability to effectively allocate attention, and it is the learners that score low on both measures that fail to extract the regularities from a vocabulary of 24 nouns. In the new items, which do not have to be memorized, the ability to allocate attention efficiently to the various task components becomes even more crucial. This is why the Reading Span task does not make an independent contribution to explaining performance in the new items. Note that although none of the correlations with

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

the Non-word Span were significant, the pattern of association was in the same direction.

Another way of conceptualizing this finding is to perform a median split of participants on Culture-Fair IQ scores, and to examine the effect of vocabulary size in the high and low IQ groups. The median Culture-Fair IQ was 24 items correct (i.e., equivalent to a standardized IQ score of 111; Cattell & Cattell, 1973), there were 28 participants below this median, and 24 above. (The remaining eight participants performed at the median and were therefore dropped from this analysis.) Performance in the old and new items as a function of vocabulary size and IQ group is depicted in Figure 2. A 2 (IQ group: high vs. low) × 3 (Vocabulary Size: 6, 12 vs. 24) ANOVA on performance in the old items yielded a main effect of vocabulary size, $F\ (2, 46) = 12.5$, $p < .001$, a main effect of IQ group, $F\ (1, 46) = 21.4$, $p < .001$, and a significant interaction between vocabulary size and IQ group, $F\ (2, 46) = 6.7$, $p < .01$, which is depicted in the upper panel of Figure 2. As we already have discussed the main effect of vocabulary size above, we will focus here on the interaction with IQ group. For old items, performance decreased markedly for the below-median IQ group with increased vocabulary size. A one-way ANOVA in the below-median IQ group showed a main effect of vocabulary size, $F\ (2, 25) = 12.3$, $p < .01$. Post hoc tests using Fisher's protected LSD showed that performance in the Size 24 group was significantly worse than performance in the Size 6 and Size 12 groups ($p < .01$). Even though performance in the Size 6 group was slightly better than performance in the Size 12 group, this difference fell short of significance ($p = .1$). In contrast, the above-median group showed strong performance regardless of vocabulary size condition, averaging 87% correct across groups. A one-way ANOVA in the above-median IQ group showed no effect of vocabulary size. The performance in the high- and low-IQ groups differed significantly only in the Size 24 condition, $t\ (14) = 4.0$, $p < .01$ (with Bonferroni correction), but not in the Size 6 and Size 12 conditions, confirming the earlier finding about the relationship between Culture-Fair IQ and performance in the old items in the Size 24 condition.

For the new items, it was crucial that a similar ANOVA yielded a main effect of IQ group, $F\ (1, 46) = 18.7$, $p < .001$, as well as a significant interaction between IQ group and vocabulary size, $F\ (2, 46) = 4.6$, $p < .05$, which is depicted in the lower panel of Figure 2. The main effect of vocabulary size was not significant, $F < 1$. One-way ANOVAs in the two IQ groups separately showed that the effect of vocabulary size fell short of significance, $F\ (2, 25) = 2.0$, $p = .1$ in the below-median IQ group, and $F\ (2, 21) = 2.9$, $p = .08$ in the above-median IQ group. Post hoc tests did not reveal any significant differences between vocabulary sizes in the below-median IQ group. In the above-median IQ group, post hoc tests showed that performance in the Size 24 group was significantly better than performance in the Size 6 group ($p < .05$), whereas performance for the Size 12 group did not differ from either Size 6 or Size 24 groups. Bonferroni corrected comparisons between the two IQ groups for each vocabulary size condition revealed that low IQ learners performed worse than high IQ learners in the Size 12 condition, $t\ (14) = 3.0$, $p < .01$, and in the Size 24 condition, $t\ (14) = 3.8$, $p < .01$, but not in the Size 6 condition. This confirms that only the high IQ learners were able to benefit from greater input variability.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

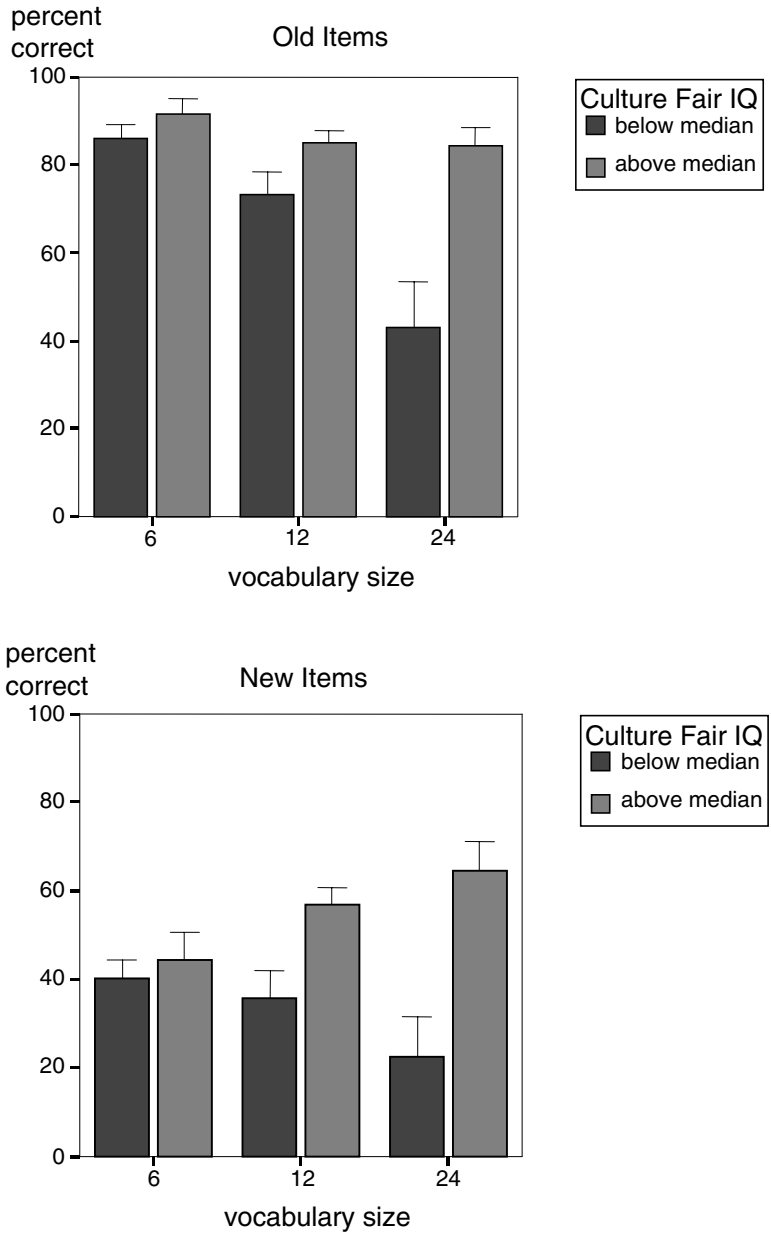


Figure 2. The percentages correct for old and new items as a function of Culture-Fair IQ and vocabulary size group. Error Bars show 1 *SEM*.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Table 5. *Mean percentage of correctly recalled nouns as a function of vocabulary size and Culture-Fair IQ*

	Vocabulary Size		
	6	12	24
Below median IQ	97.0 (6.7)	80.8 (24.5)	52.4 (18.9)
Above median IQ	87.0 (13.9)	97.2 (6.8)	79.2 (14.9)

Note: Standard deviations are in parentheses.

Vocabulary test performance

To determine how well the participants were able to remember the Russian words, we counted the percentage of correctly recalled items in the Vocabulary Test administered at the last session. Because many participants had problems with pronunciation of the unfamiliar Russian phonology, correct recall was coded liberally: if the initial phoneme as well as the basic shape of the word stem resembled the original Russian word, the item was classified as correctly recalled. Table 5 summarizes the percentage of correctly recalled nouns for each of the three vocabulary size conditions for learners with above- and below-median Culture-Fair IQ. A 3 (Vocabulary Size: 6, 12, 24) × 2 (Culture-Fair IQ: below vs. above) mixed-type ANOVA yielded a main effect of vocabulary size, $F\ (2, 46) = 13.6$, $p < .001$, a main effect of Culture-Fair IQ, $F\ (1, 46) = 6.1$, $p < .05$, as well as a significant interaction, $F\ (2, 46) = 6.4$, $p < .01$. Post hoc tests using Fisher's protected LSD showed that participants in the Size 24 group performed worse than participants in the Size 6 and Size 12 groups ($p < .001$).

As expected, vocabulary recall was better the smaller the number of nouns in the learning set that had to be remembered. Apparently, the participants in the Size 6 and Size 12 conditions had no problems recalling the nouns, but in the Size 24 group, participants with above-median IQ recalled significantly more nouns than participants with below average IQ ($p < .01$).

Error analyses

In the final part of this section we will present a more detailed analysis of the kinds of errors committed in the different vocabulary size conditions to understand what aspects of learning are affected by vocabulary size and by the individual differences in executive functioning. We classified the errors according to the following categories, which correspond to the task components described earlier: (a) noun repetition errors were responses in which participants produced the wrong noun regardless of whether this noun was used with the right preposition or whether it was correctly inflected. These errors indicated that the participant was not able to reproduce the basic phonological shape of the target noun. (b) Preposition errors were responses in which participants repeated the phrase with the noun introduced in the nominative case (e.g., they produced *eto zabor* instead of *k zaboru*) or else they produced the wrong preposition (e.g., they produced

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

ot zabora instead of *k zaboru*). These errors indicated that the participant was not able to select the appropriate preposition associated with the question and the direction of movement in the picture, thus indicating basic difficulties in learning to produce the correct prepositions in appropriate contexts. The overwhelming majority of these errors involved repetition of the nominative form. (c) Within-paradigm errors were responses in which participants selected the alternative case marker for a particular noun, for example, producing *ot zaboru* instead of *ot zabora* or *k zabora* instead of *k zaboru*. These errors indicated that the participant had learned the gender of the noun but was confusing the dative and genitive case markers. That is, the participant seemed to have learned how to inflect the particular noun, but was incorrect in selecting the form appropriate for the required case. (d) Inflection errors (not otherwise specified) were responses in which participants produced the appropriate preposition, followed by the noun in the nominative case, or else inflected with any inflection besides the genitive or dative marker appropriate for that particular noun. Unfortunately, due to the existence of inflectional syncretism in the Russian paradigm (i.e., the inflection *-a* is ambiguous between feminine-nominative and masculine-genitive) it was essentially impossible for us to characterize the source of many of these errors. For example, if a participant produced *ot zvezda* instead of *ot zvezdy*, we could not ascertain whether the *-a* inflection was an instance of feminine-nominative (i.e., correct gender, wrong case) or masculine-genitive (i.e., wrong gender, correct case). If a participant produced *ot zabor* instead of *ot zabora*, we could not determine the extent of their knowledge of the gender classification of the noun.

Figure 3 shows the distribution of error types in the three vocabulary size conditions for participants with below-median and above-median Culture-Fair IQ scores in the old items. The main observation here is that the larger error rates in the below-median IQ group are mainly due to a failure to select the correct preposition. It appears that, when learners are confronted with variability in the input, low attentional capacity undermines the learning process already at a very basic level: these participants had considerable difficulties in associating the direction of the movement, and the associated question, with the correct preposition. Note that each training session devoted an entire block of trials (Block 3) to testing participants' ability to distinguish the dialogues for dative and genitive cases. Yet still, some of the participants failed to acquire the very basic associations between prepositions and depicted events, as measured in their productions. What are the repercussions of this basic learning difficulty? Figure 4, which presents error types for new items, shows that in the Size 24 condition, below-median IQ participants still exhibited a considerable proportion of preposition errors. In contrast, the above-median IQ participants had no problems in learning the prepositions, and were able to focus their attention on selecting the appropriate case markers.[2]

DISCUSSION

The present work addressed two main questions: first, how does the frequency of patterns in the input affect the learning of complex inflectional systems in adult L2

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

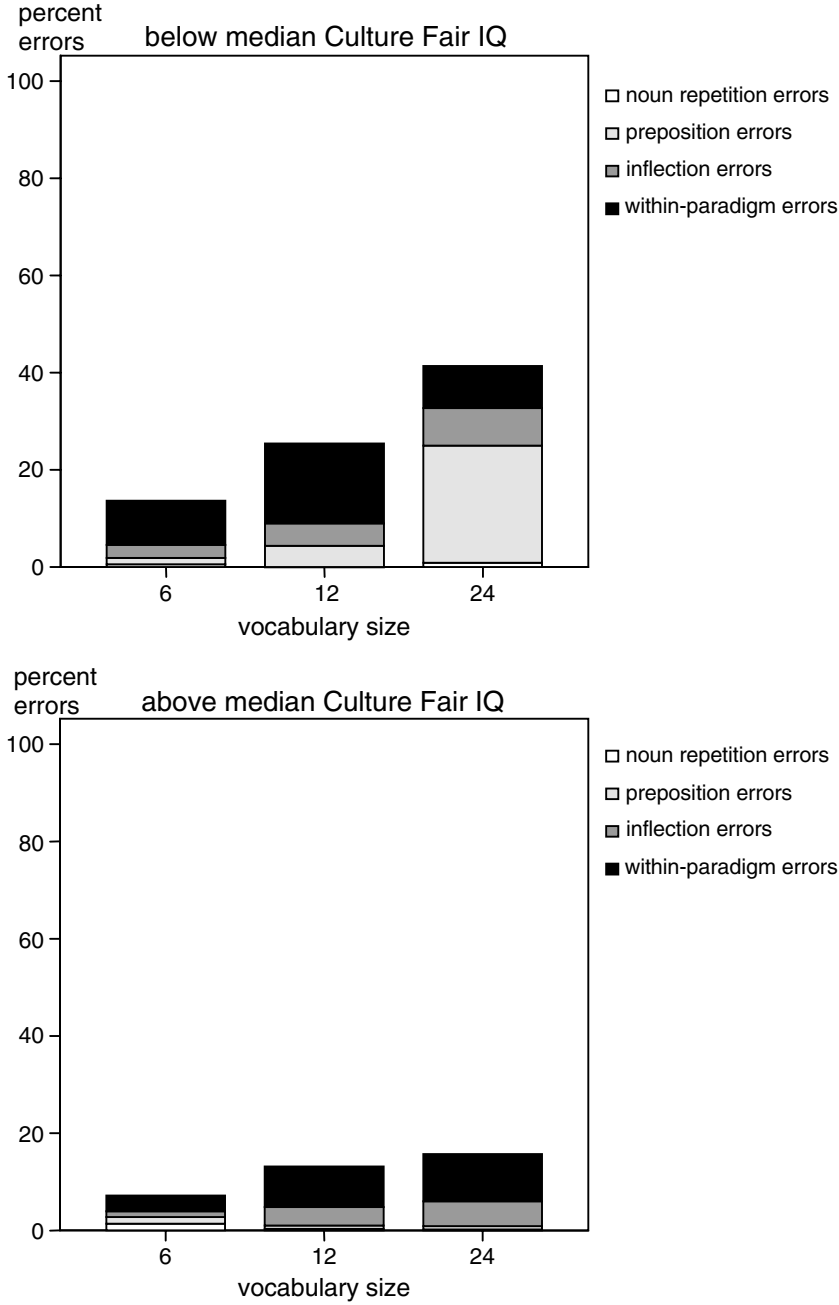


Figure 3. The distribution of error types for old items for participants with below-median and above-median Culture-Fair IQ scores as a function of vocabulary size group.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

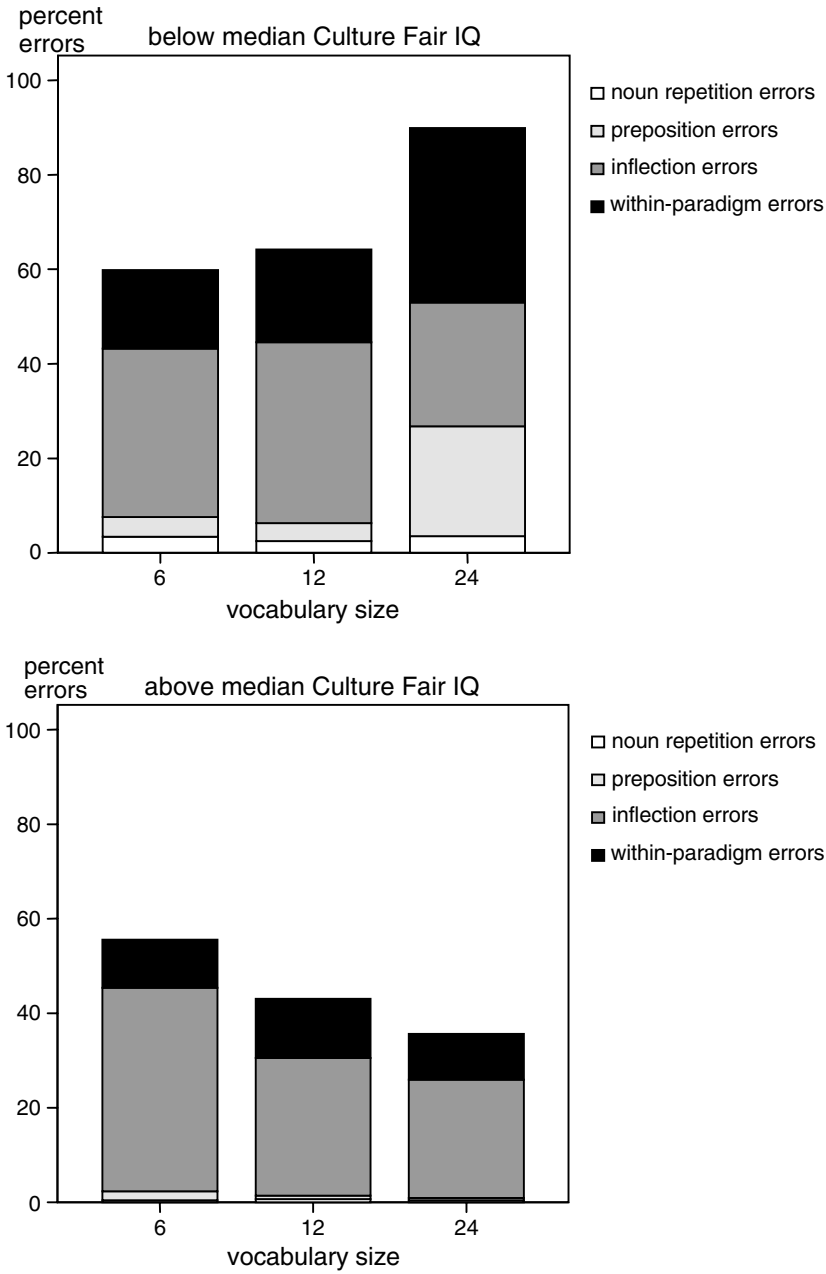


Figure 4. The distribution of error types for new items for participants with below-median and above-median Culture-Fair IQ scores as a function of vocabulary size group.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

learners? In the present study, we trained adult L2 learners extensively on a subset of Russian noun morphology, and tested their ability to inflect the nouns presented during training as well as new nouns as a measure of generalization. During training, the overall frequency of the inflectional patterns was held constant in the input (i.e., all participants received the same number of learning trials), which allowed us to test the effect of variation in type frequency on learning of Russian case marking patterns. Our hypothesis was that increased type frequency would facilitate generalization by forcing learners to construct categories that were general enough to provide coverage for all items instantiating the inflectional patterns in the input. Our second question was whether individual differences in basic information processing capabilities would affect learners' ability to extract the inflectional patterns present in the input. Although studies of individual differences (e.g., Kane & Engle, 2002; Miyake, 2001; Miyake, Friedman, & Emerson, 2000; Price, 2004) are receiving increased attention in education and in cognitively oriented basic research, relatively little is known about the mechanisms by which individual differences in cognitive capabilities influence language learning in general, and morphology learning in particular.

Our main finding was that adult L2 learners varied greatly in their ability to benefit from richer input in the context of learning a complex inflectional paradigm: although a larger training vocabulary was crucial for generalization, it could only be acquired and utilized by learners who possessed sufficient attentional resources to handle the rich input. High IQ learners acquired more vocabulary in the Size 24 group (in comparison to the low IQ group), which resulted in their having access to a broader database for extracting inflectional regularities. Thus, in adult L2 learners, the effect of a critical mass on morphology learning (Marchman & Bates, 1994) appeared to be mediated by the learner's ability to expend cognitive effort and to allocate attention to the various components of the learning task. This is an important addition to the results obtained by Gomez (2002), who showed that generalization of nonadjacent dependencies to novel items in an artificial language was only possible in learners exposed to a large enough vocabulary. Our study demonstrates that when it comes to learning a morphologically rich language, the situation is considerably more complex. Learners are faced with variability on various levels, and have to solve multiple categorization problems such as mapping morphosyntactic patterns to differences in meaning, categorizing words according to grammatical categories, discovering the domain of application for different inflectional paradigms, detecting allomorphic variation, and learning to ignore it in syntactic categorization. Handling all these categorization tasks is complicated by the unfamiliar phonotactics of the L2, which impairs the detection of phonological characteristics associated with the various morphosyntactic patterns, and last but not least, by the articulatory difficulties associated with producing phonotactically unfamiliar words.

In the present task, learning the complex Russian inflectional patterns involved a number of components, and we discovered that Culture-Fair Nonverbal IQ strongly predicted learners' success. The most basic task component involved learning to associate one of two Russian prepositions with a question about, and depicted direction of, an elephant's movement. Importantly, learners with low Culture-Fair IQ scores, when faced with a large amount of variability in the input,

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

often failed to register even the most basic dependencies between the prepositions and the questions/pictures, and that failure affected both the old and the new items (see "preposition" errors in Figures 3 and 4). The next task component involved learning to associate the Russian prepositions with the case-marking vowel endings of the nouns. Case marking in Russian, however, depends not only on the direction of the movement expressed by the preposition, but also on the gender of the noun. Thus, to select the correct case suffix, learners had to first identify noun gender. In our study, we used masculine and feminine nouns belonging to the so-called first and second Russian declensions. Gender category membership was always transparently marked in the nominative case to facilitate the learning of the gender categories (Brooks et al., 1993; Kempe & Brooks, 2001). That is, all feminine nouns ended in *-a* and all masculine nouns ended in consonants. Unfortunately, a limitation of the present study is that we simply do not know how well the underlying gender categories were learned, due to the presence of inflectional syncretism precluding fine-grained analysis of error types. A better design for future studies would be to include other measures of gender learning (e.g., adjective–noun agreement) that are not affected by inflectional syncretism.

The final task component consisted of selecting the right suffix for the target case within the gender category. Overall, the "within-paradigm" errors (see Figures 3 and 4) were more prevalent in below-median IQ group than in the above-median IQ group. Thus, the low-IQ participants typically exhibited greater difficulties in learning the prepositions, and in learning to associate the right case suffixes with the appropriate contexts within each gender category. Further work is needed to determine the extent to which the observed failures to learn in the low-IQ and small set size groups were related to particular aspects of the linguistic system being trained, and whether similar findings would obtain in adults exposed to a different language.

Although Culture-Fair IQ turned out to be the strongest predictor of individual differences in the observed benefits of increased type variability in learning Russian case-marking patterns, Reading Span scores also correlated with success in learning for the vocabulary size 24 condition. Regression analyses showed that, for new items, Reading Span did not predict significant unique variance in performance after the effect of Culture-Fair IQ was removed. This makes it likely that the contribution of Reading Span to success in generalizing Russian case marking involved executive components in common with the Culture-Fair Test. The Non-word Span Test of phonological memory, although found to be moderately correlated with the Culture-Fair Test, fell short of significance in predicting learning success for the vocabulary size 24 condition. These results, taken together, suggest that there might be considerable overlap across our three predictor tasks in the variance components predictive of successful grammar learning in individuals exposed to highly diverse vocabulary. We speculate that these tasks differ with respect to at least two dimensions: first, Culture-Fair, Reading Span, and Non-word Span tests differ with respect to the involvement of language production. Thus, Culture-Fair and Non-word Span do not require explicit language production, whereas Reading Span does. Second, performance on the Reading Span task may reflect the disruptive effects of articulatory suppression on

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

verbal memory (Hupet, Desmette, & Schelstraete, 1997). These differences in task requirements may explain the lack of a correlation observed between the Reading Span task and the other two predictor tasks. In addition, the three tasks differ in the extent to which focused modulation of attention is required. Extracting complex morphosyntactic regularities from the input in the way described above requires executive involvement, which is a central component of the Reading Span task and, to an even greater extent, the Culture-Fair IQ test. Clearly, additional studies are needed with larger samples to more fully tease apart the task components of these three individual difference measures to further specify the relationships between task components and morphology acquisition. Furthermore, given the findings of Fry and Hale (1996, 2000) that individual differences in processing speed are strongly correlated with working memory capacity (and indirectly related to fluid intelligence), these future studies should include measures of processing speed to examine its contribution to successful L2 learning in adults.

The finding that Culture-Fair IQ mediated success in language learning in adults confirms previous work showing moderate correlations between IQ and language learning aptitude (e.g., Grigorenko et al., 2000; Sasaki, 1996; Skehan, 1989). Taken together, the results of these studies resonate with a growing literature arguing for a broader role of explicit processing in L2 acquisition than in L1 acquisition (e.g., N. Ellis, 1994; Norris & Ortega, 2000). That is, when implicit and explicit language instructional methods are compared, providing explicit instruction is more beneficial for adult learners (see Norris & Ortega, 2000, for a meta-analytic review). Following Schmidt (1990) and Robinson (1995), Ellis (2002) has advanced a "noticing" hypothesis that attention must be paid to an aspect of the linguistic stimulus before a mental representation of that aspect can be formed. Once that representation is firmly established, its strength can be incremented through implicit processing. He has further suggested that certain complex associations, such as long-distance grammatical dependencies, cannot be learned by adult L2 learners without conscious explicit learning and hypothesis testing. "Noticing lays out the problem. Consciousness-raising can speed its solution" (Ellis, 2002, p. 175). Taraban (2004) provides supporting evidence that explicit processing is strongly facilitative, and possibly required, for adults' successful learning of an artificial language with gender categories and case marking patterns that were similarly structured to the Russian categories used in the present study (see also Brooks et al., 1993). Frenck-Mestre et al. (2004) have further argued, based on behavioral and ERP data, that experience with grammatical gender in an L1 is probably necessary for full mastery of a gender system in L2.

Thus, there is an emerging consensus that attentional mechanisms, executive supervisory systems and meta-linguistic awareness play crucial roles in adult learning of complex grammatical dependencies and regularities. This conclusion is also consistent with work in behavioral genetics, which suggests that with age there are increasingly strong links between general cognitive abilities (such as working memory capacity and fluid intelligence) and linguistic abilities (Dionne et al., 2003). Our study adds an important contribution by showing that the learning of complex morphosyntactic paradigms by adult L2 learners is contingent upon both sufficient variability in the input, and the ability to utilize this rich input through

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

allocation of attentional resources to the various components of the language learning task.

APPENDIX A

Training and testing vocabularies

Training		
Size 6	Size 12	Size 24
Zabor, masc (fence)	Zabor, masc (fence)	Zabor, masc (fence)
Samoljot, masc (airplane)	Samoljot, masc (airplane)	Samoljot, masc (airplane)
Mjachik, masc (ball)	Mjachik, masc (ball)	Mjachik, masc (ball)
Zvezda, fem (star)	Kuvshin, masc (pitcher)	Kuvshin, masc (pitcher)
Sosiska, fem (sausage)	Vinograd, masc (grapes)	Vinograd, masc (grapes)
Metla, fem (broom)	Banan, masc (banana)	Banan, masc (banana)
	Zvezda, fem (star)	Sharik, masc (balloon)
	Sosiska, fem (sausage)	Limon, masc (lemon)
	Metla, fem (broom)	Poezd, masc (train)
	Kniga, fem (book)	Stakan, masc (glass)
	Vershina, fem (summit)	Chajnik, masc (teapot)
	Bochka, fem (barrel)	Budil'nik, masc (clock)
		Zvezda, fem (star)
		Sosiska, fem (sausage)
		Metla, fem (broom)
		Kniga, fem (book)
		Vershina, fem (summit)
		Bochka, fem (barrel)
		Shljapa, fem (hat)
		Luna, fem (moon)
		Pila, fem (saw)
		Skripka, fem (violin)
		Chashka, fem (cup)
		Babochka, fem (butterfly)
Testing (All Conditions)		
Banan, masc (banana)		
Kostjum, masc (suit)		
Baraban, masc (drum)		
Limon, masc (lemon)		
Paket, masc (parcel)		
Telefon, masc (telephone)		
Bochka, fem (barrel)		
Butylka, fem (bottle)		
Luna, fem (moon)		
Polka, fem (shelf)		
Kepka, fem (cap)		
Tarelka, fem (plate)		

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

ACKNOWLEDGMENTS

This project was supported by a grant from *Language Learning—A Journal of Research in Language Studies*. We thank Nicholette Apap, Jennie Chang, Kermeel Charles, and Ivan Scott C. M. Lee for their assistance in conducting the experiment, and Rebecca Gomez, Anat Ninio, and Irina Sekerina for their comments on the manuscript.

NOTES

1. The use of production data as our dependent variable unfortunately conflates knowledge of the appropriate forms with participants' ability to produce them. We chose not to use a grammaticality judgment task (i.e., presenting participants with grammatical and ungrammatical strings for their evaluation) because of concerns that it would be difficult for a native speaker to produce ungrammatical inflected forms without any special or abnormal emphasis, and because production of L2 forms is a more ecologically valid task.
2. We also conducted analyses of the errors as a function of noun gender and case. These analyses revealed that learners in general had more difficulties with the declension of feminine nouns, which we suspect is due to the greater phonological similarity between the feminine genitive *-y/-i* and dative *-e* suffixes in contrast to masculine *-a* versus *-u*, and allomorphic variation in the feminine genitive. However, because these analyses are quite complex, and not crucial to the central point of this paper, they are not presented here but can be made available upon request.

REFERENCES

Adams, A.-M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language and Communication Disorders, 35*, 95–116.

Baddeley, A., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

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*, 507–584.

Braine, M. D. S. (1987). What is learned in acquiring word classes—A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Erlbaum.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J., Catalano, L., et al. (1990). Exploring language acquisition in children through the use of a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29*, 591–610.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76–95.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425–255.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

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., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instrumentation, and Computation, 25*, 257–271.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*, 45–75.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*, 394–412.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*, 257–303.

Ellis, N. (Ed.). (1994). *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N., & Schmidt, R. (1998). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition, 19*, 145–171.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Frenck-Mestre, C. A., Foucart, A., & Caetano-Nunes, E. (2004). Accessing grammatical gender in a second language: Native language influences. *Abstracts of the Psychonomic Society, 45th Annual Meeting, 9*, 47.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*, 1–34.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gesi Blanchard, A. T. (1998). Transfer effects of first language proficiency on second language reading. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 291–314). Mahwah, NJ: Erlbaum.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

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

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*, 390–405.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*, 220–225.

Hupet, M., Desmette, D., & Schelstraete, M.-A. (1997). What does Daneman and Carpenter's Reading Span really measure? *Perceptual and Motor Skills, 84*, 603–608.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.

Kane M. J., & Engle, R. A. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9*, 637–671.

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*, 221–256.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press

MacWhinney, B. (1992). Transfer and competition in second language learning. In: Harris, R. J. (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: North-Holland.
Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.
Miyake, A. (2001). Individual differences in working memory: Introduction to special section. *Journal of Experimental Psychology: General, 130*, 163–168.
Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In Healy, A. F., Bourne, L. E., Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.
Miyake, A., Friedman, N. P., & Emerson, M. J. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cogntive Psychology, 41*, 49–100.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.
Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*, 463–490.
Price, L. (2004). Individual differences in learning: Cognitive control, cognitive style, and learning style. *Journal of Educational Psychology, 24*, 681–698.
Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*, 283–331.
Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. Baltimore, MD: Peter Lang.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Psycholinguistics, 11*, 129–158.
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*, 4–27.
Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.
Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language, 51*, 202–216.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Winke, P. M. (2005). *Individual differences in adult Chinese second language acquisition: The relationship among aptitude, memory and strategies for learning*. Unpublished PhD dissertation, Georgetown University.

https://doi.org/10.1017/S0142716406060243 Published online by Cambridge University Press