Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

The role of distributional factors in learning and generalising affixal plural inflection: An artificial language study

Michael Nevat, Michael T. Ullman, Zohar Eviatar & Tali Bitan

LANGUAGE, COGNITION AND NEUROSCIENCE
2018, VOL. 33, NO. 9, 1184–1204
https://doi.org/10.1080/23273798.2018.1465187

REGULAR ARTICLE

The role of distributional factors in learning and generalising affixal plural inflection: An artificial language study

Michael Nevat[a], Michael T. Ullman[b], Zohar Eviatar[c] and Tali Bitan[a,c,d]

[a]The Language and Brain Plasticity Lab, University of Haifa, Haifa, Israel; [b]Department of Neuroscience, Georgetown University, Washington DC, USA; [c]Department of Psychology, IIPDM, University of Haifa, Haifa, Israel; [d]Department of Speech Language Pathology, University of Toronto, Toronto, Canada

ABSTRACT
Inflectional morphology has been intensively studied as a model of language productivity. However, little is known about how properties of the input affect the emergence of productive affixation. We examined effects of three factors on the learning and generalisation of plural suffixation by adults in an artificial language: affix type frequency (the number of words receiving an affix), affix predictability (based on phonological cues in the stem), and diversity (the number of distinct phonological cues predicting an affix). Higher type frequency and predictability facilitated the acquisition of trained inflections. Type frequency contributed to participants' inflections of untrained words early during learning, while reliance on diversity emerged gradually, alongside knowledge of phonological cues. Diversity as well as type frequency contributed to the emergence of default-like inflections, including minority defaults. The results elucidate the role of affix diversity and its interaction with other factors in the emergence of productive linguistic processes.

ARTICLE HISTORY
Received 13 May 2017
Accepted 10 April 2018

KEYWORDS
Inflection; artificial language; generalisation; diversity

Introduction

Morphological inflection has long been examined as a model for investigating productive aspects of language (e.g. Dabrowska, 2001; Marchman, 1997; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Prasada & Pinker, 1993; Rumelhart & McClelland, 1986). However, to date there is no consensus regarding the extent to which speakers' productivity in using inflections depends on the statistical properties of the input. The current study examines the effect of such "distributional factors," namely, relative frequencies and other probabilities, on the learning and generalisation of morphological affixal inflections in adults learning a novel language.

Various factors have been found to affect the generalisation of inflections to unfamiliar words in participants' first language. One important factor is inflection type frequency (i.e. the number of words taking each inflection) (Dabrowska & Szczerbiński, 2006; Hsieh, Leonard, & Swanson, 1999; Köpcke, 1998; Köpcke & Wecker, 2017; Marchman, 1997; Maslen, Theakston, Lieven, & Tomasello, 2004). A second factor is the inflection's predictability (i.e. the degree to which the inflection can be predicted from phonological or other cues in the stem. e.g. Albright & Hayes, 2003; Hartshorne & Ullman, 2006; Köpcke & Wecker, 2017; Marchman, 1997).

Finally, and crucially for the present study, a third factor, related to the presence of phonological cues in stems, is the inflection's phonological "diversity," which refers to the number of distinct phonological cues present in the stems taking each inflection. There is some evidence for the effect of inflections' phonological diversity on generalisation in participants' first language. This factor has been found to facilitate generalisation to unfamiliar words in Polish speaking children (e.g. Dabrowska, 2004; Dabrowska & Szczerbiński, 2006), and the effect of diversity on generalisation appears to increase with age during childhood (Dabrowska & Szczerbiński, 2006).

Some scholars argue against reliance on such statistical information as being the sole basis for productivity with inflections (e.g. Berent, Pinker, & Shimron, 1999; Marcus et al., 1995; Pinker, 1991; Prasada & Pinker, 1993). These scholars point to evidence suggesting that the application of regular inflections (such as the suffix –ed in English past tense) does not depend on their type frequency or their predictability. Regular inflection, therefore, serves as a "default" inflection (the inflection to be applied in a variety of conditions in which phonological similarity to familiar words does not provide sufficient information), even when it is not the most common

CONTACT Tali Bitan tbitan@research.haifa.ac.il
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2018.1465187

inflection (Marcus et al., 1995). However, the existence of such "minority default inflections" may still be explained by reliance on other types of statistical information. It is conceivable that an inflection with high phonological diversity (namely, an inflection which applies to stems that are highly diverse in terms of the phonological cues they contain) might be perceived as "cue-independent," and therefore emerge as a default inflection regardless of its frequency. Marcus et al. (1995) confirm that this may indeed be the case with the German plural affix -s (p. 245). Understanding the effect of phonological diversity on the learning and generalisation of inflections could therefore shed light on the emergence of default-like preferences for certain inflections. Moreover, studies on other aspects of language learning (i.e. other than morphological inflection), as well as non-linguistic learning, point to the relevance of diversity for learning and generalisation, as will be described in the following paragraph.

One example for the effect of stimulus diversity on generalisation in a linguistic context comes from an artificial language learning study in which adults were trained on identifying valid syntactic structures, and were then tested on generalizing their knowledge to untrained items (Reeder, Newport, & Aslin, 2013). The results showed better generalisation by participants trained on highly diverse stimuli compared to those trained on a smaller number of distinct items, despite an identical number of trials. The benefit of a diverse training set for broader generalisation was also found in infants learning an artificial language (Gerken, 2006) and adults learning to read in an artificial script (Aduan-Mansour & Bitan, 2017). Finally, a non-linguistic category learning study in adults (Hahn, Bailey, & Elvin, 2005) showed that within-category diversity facilitates generalisation. Thus, diversity seems to improve the learning and especially generalisation in a range of linguistic and non-linguistic domains.

To date most research on the effects of the distributional factors examined here (inflection type frequency, predictability, and phonological diversity) on the processing of morphological inflections has focused on first language (L1), particularly in children. Second language (L2) learning may rely on somewhat different mechanisms, due to participants' linguistic experience and age of acquisition (Babcock, Stowe, Maloof, Brovetto, & Ullman, 2012; Hudson Kam & Newport, 2005, 2009; Ramscar & Gitcho, 2007). More specifically, various empirical findings and models have suggested that L2 speakers tend to store inflected forms more than L1 speakers, and are less sensitive to the morphological structure of inflected words, especially when the L2 is acquired in adulthood or is less proficient (Babcock et al., 2012; Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Morgan-Short and Ullman, 2012; Silva & Clahsen, 2008; Ullman, 2001, 2015). However, evidence also suggests that, with enough experience with the L2, affixal inflected forms may increasingly depend on rule-governed compositional processes (Babcock et al., 2012; Morgan-Short and Ullman, 2012; Ullman, 2001, 2015). Nevertheless, the evidence is still sparse, and storage vs. composition differences between L1 and L2 in the processing of morphological inflections are still a matter of debate (Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Feldman, Kostić, Basnight-Brown, Filipović Đurđević, & Pastizzo, 2010; Tolentino & Tokowicz, 2011). Here we focus on distributional factors in the learning of affixal morphology rather than this debate, though we return to it briefly in the Discussion.

One of the few studies that have examined distributional factors affecting inflections in L2, compared L1 and L2 German speaking children applying German plural inflections to non-words (Köpcke & Wecker, 2017). The results showed reliance on both the inflections' type frequency, and their predictability in generalisation to new words, in both groups (albeit with different developmental trajectories). In adults, Kempe and Brooks (2008) found that English L1 speakers learning the Russian case marking system rely on the predictability of the suffixes given the nouns' grammatical gender. Finally, results of studies using artificial languages also indicate that inflections' type frequency affects learning (Ellis & Schmidt, 1997; 1998), as well as generalisation (Bybee & Newman, 1995). However, the effects of inflections' phonological diversity on learning and generalisation in a second language or an artificial language have not been studied. The current study therefore sought to examine how the effects of affixal phonological diversity evolve and interact with those of affix type frequency and affix predictability, and particularly how these factors may contribute to the emergence of minority default inflections.

The current study

In order to examine the effects of distributional factors on the learning and generalisation of morphological inflection we designed an artificial language. Artificial language paradigms are particularly well suited for examining learning and generalisation because one can tightly control the amount and type of exposure, including manipulating factors of interest in the input. Artificial linguistic paradigms have the added advantage that, likely because they are small, they can generally be learned to reasonably high proficiency over the course of hours to days. Hence, despite concerns regarding their ecological validity, because they do not reflect the full complexity of natural languages learned in natural settings, artificial languages have been widely used in the investigation of learning both vocabulary (Tamminen, Davis, Merkx, & Rastle, 2012) and grammar (e.g. Ellis & Schmidt, 1997; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Importantly, although researchers have used artificial language paradigms as models of either first or second language acquisition, in the present study we interpret the learning and generalisation of the artificial language as a model of second language learning. This is because in this study learning occurs in adulthood (when participants have already learned the first language), and moreover, as in a second language, the artificial language involved new words referring to familiar items (e.g. apple; see Methods). Indeed, results of studies using artificial languages indicate that participants' performance on the artificial language are positively correlated with measures of acquisition of a natural L2 (Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, 2016). Note that training on an artificial language can result in native-like brain activity patterns (Morgan-Short et al., 2012), suggesting that results from artificial languages can indeed generalize to (either first or second) natural languages.

The artificial language in the current study consisted of nouns referring to familiar objects, and their plural affixal inflections. Plural inflection of nouns was examined because nouns lend themselves to experimental designs in which novel nouns are associated with objects (e.g. by simultaneously presenting the noun and an image of its referent). Plural inflection of nouns has indeed been investigated in multiple artificial language studies (Bybee & Newman, 1995; Ellis & Schmidt, 1997), as well as in the processing of natural first (Berko, 1958; Köpcke & Wecker, 2017; Laaha, 2011; Marcus et al., 1995; Vender, Mantione, Savazzi, Delfitto, & Melloni, 2017) and second language (Jia, 2003; Kaivapalu & Martin, 2007; Köpcke & Wecker, 2017). Indeed, plural inflection tends to be acquired relatively quickly, as compared to verbal inflection (Brown, 1973; Hsieh et al., 1999; Wood, Kouider, & Carey, 2009) – likely in part due to the lexical/semantic reality of nominal plural inflection (Pinker, 1984) – facilitating the examination of the course of learning. Thus, affixal plural inflection is a reasonable target for the examination of the acquisition of affixal inflection, although it remains unclear to what extent findings from plural affixal inflection can be generalized to other aspects of inflection (see Discussion).

Three factors were manipulated in the current study: affix type frequency, affix predictability based on the phonological (rime) cues, and the affix phonological diversity, based on the rime cues. Three groups of adults were trained, each group on a different list of trained words. Affix type frequency was manipulated by varying the type frequencies of the suffixes within each group. The (relative) type frequency of a suffix is given by n_{suffix}/N where n_{suffix} is the number of different words taking the suffix, and N is the total number of trained words. Affix predictability (given a cue contained in the stem) is defined as $n_{suffix\cap cue}/n_{cue}$, where $n_{suffix\cap cue}$ is the number of words containing the cue and taking the suffix, and n_{cue} is the total number of words containing the cue. Affix predictability differed between groups as well as between affixes within group. Finally, affix phonological diversity was defined as $\sum_{cues}(n_{suffix\cap cue}/n_{cue})/N_{cues} = \sum_{cues}$ (Predicability (suffix, cue))$/N_{cues}$, where $\sum_{cues}()$ indicates a sum across all distinct cues, and N_{cues} is the total number of distinct cues. Affix diversity was also manipulated both between and within group. In addition to the learning of trained items, we examined generalisation to untrained items, of two types. Untrained items containing phonological cues were included to test the reliance on affix predictability. Additionally, untrained items without phonological cues were included to examine the factors that contribute to the emergence of default inflections.

Based on results of previous studies, we predicted that high affix type frequency and high affix predictability would facilitate learning trained items and generalisation to untrained items containing phonological cues. Importantly, based on finding showing the effect of diversity on generalisation of inflections in L1, as well as in other domains, we expected to find an effect of affix phonological diversity on generalisation to untrained items without phonological cues. Specifically, we predicted that affixes with high type frequency and affixes with high phonological diversity might serve as "default inflections" for words which are not phonologically similar to trained words. Finally, we expected to find earlier effects of affix type-frequency on learning and generalisation, as compared to the other two factors. This is because reliance on both affix predictability and affix phonological diversity depends on the acquisition of knowledge about phonological cues, which is expected to emerge more gradually.

Methods

Participants

Three groups of 18 participants each (12 women in each group), ages 21 to 43.5 (mean 25.77) were recruited for this study. All participants were native Hebrew speakers and spoke at least one other language (English) as a second language. All participants reported being right-handed, had no known psychiatric, neurodevelopmental or neurological disorders, and had normal or corrected hearing and vision. Participants were randomly assigned to the three groups, with equal assignment of males and females across groups. Each group trained on a somewhat different list of words, henceforth referred to as lists A, B and C. Lists of trained words were designed so that lists A and B differ in the distribution of phonological diversity across affixes, and lists A and C differ in terms of the affixal phonological predictabilities, while other factors were designed to be as similar as possible across groups.

Materials

All participants learned the plural inflections of 48 aurally presented nouns in a novel language. All words consisted of two syllables, with a common structure (CVCVC), in their singular form. Plural forms were obtained by applying one of 5 possible (VC) suffixes to the singular form. The "high-frequency" suffix was applied to half of the trained words (24), the "medium frequency" suffix was applied to one quarter of the trained words (12), and three "low-frequency" suffixes were each applied to one twelfth (4) of the trained words. Pairings of trained words and suffixes were generally determined by words' rimes (e.g. words ending with /oz/ took the high-frequency suffix "–an", and thus the plural for "tuvoz" was "tuvozan"; see Table 1), although participants were not informed of these cues. However, lists A and B contained exceptions to this rule, i.e. words containing "inconsistent rime cues." Each of these words was inflected with a suffix that differed from the suffix taken by the majority of words that it rhymed with. For example, the word "shalod" (from list A), was inflected with the high-frequency suffix /-an/ although it rhymed with "napod" and "resod", which were inflected with the medium-frequency suffix "-esh" (see Table 1). These words with inconsistent rime cues were included in order to manipulate the predictability of affixes given these phonological cues. Words containing inconsistent rime cues received the high-frequency suffix in list A, and the medium-frequency suffix in list B. None of the words in list C contained inconsistent rime cues, rendering rime cues for this group deterministic.

Affixal diversity was manipulated independently from both affix type-frequency and from affix phonological predictability. This was achieved by controlling the number of distinct rime cues found among words taking each suffix. Thus, words taking the high-frequency suffix in list A contained 9 distinct rime cues (/oz/, /ig/, /ul/, /od/, /iv/, /un/, /us/, /il/ or /om/; see Table 1), rendering it more diverse than the medium-frequency suffix in list A, which consisted of only 3 distinct rime cues (/od/, /iv/, /un/). On the other hand, for words in list B, the medium-frequency suffix was more diverse, consisting of words with 9 distinct cues, whereas the high-frequency suffix consisted of words with only 3 distinct cues. None of the words taking low frequency suffixes in any of the lists rhymed with each other (hence each word had a different rime cue). See Table 1 for the design of the lists of trained words, and the third formula above for the calculation of diversity in the current study.

Auditory stimuli were recorded by a female native speaker of Hebrew at 44,100 Hz and 16 bits per sample. Vowels and consonants were pronounced as in Hebrew. Stress was always on the last syllable, and was shifted to the suffix in plural forms, as in Hebrew. However, unlike some Hebrew words, the vowels in the stem remained unchanged despite the shift in stress. The design of the inflectional system in the current study resembles natural languages in a number of respects. The existence of multiple suffixes with varying type frequencies is found in the plural inflection system of nouns in German (Laaha, 2011; Marcus et al., 1995). Phonological cues at word ending positions predicting inflections are found in the plural inflection of nouns in Hebrew (Berent et al., 1999; Ravid et al., 2008) and German (Laaha, 2011). However, the use of an artificial language in the present enabled the full control over other properties. Hence, the first three phonemes of each word (CVC) were designed to ensure they would not provide any cue to the inflection. More importantly, the independent manipulation of affix type frequency, phonological predictability and affixal diversity was aimed at separating the effects of each factor on learning and on the emergence of a default-like preference in generalisation.

In addition to the 48 items that participants were trained on, they were tested (at the end of the first and last sessions) on the production of plural forms of 96 untrained words in each test (two lists in total) (See Table 2 and appendix B). Seventy three of the untrained words in each test rhymed with trained stems (referred to henceforth as "untrained words containing rime cues)". These served to assess participants' reliance on phonological predictability. It should be noted that for groups A and B, 24 of these 73 words contained "ambiguous rime cues" (rime cues which predicted two suffixes). For all groups the list included 23 untrained words that did not rhyme with any of the trained stems ("untrained words *without* rime cues"). These were included to examine whether participants had developed a general preference for any of the suffixes.

Table 1. Design of trained items.

	Trained-item list		
	A	B	C
	Diversity: 0.283	Diversity: 0.148	Diversity: 0.375
High frequency inflection (24 words) Suffix: "an"	Words containing "consistent rime cues" Predictability: 1.0 3 families of 6 words each: nifoz nishig tizul tuvoz posig shuzul kufoz bolig mupul laloz dedjig suful refoz rekig tedjul gishoz givig bikul Words containing "inconsistent rime cues" • 3 words rhyming with words that take the medium frequency inflection (shalod, gukiv, gitun). Predictability: 0.2. • 3 words rhyming with words taking low frequency inflections (kunus, gomil pakom). Predictability: 0.5.	Words containing "consistent rime cues" Predictability: 0.89 3 families of 8 words each: nifoz nishig tizul tuvoz posig shuzul kufoz bolig mupul laloz dedjig suful refoz rekig tedjul gishoz givig bikul savoz zobig ladul refoz kinig pokul	Words containing "consistent rime cues" Predictability: 1.0 9 families of 2–3 words each: tuvoz bolig mupul basish laloz dedjig suful safish gishoz rekig bikul tazos napod tepiv koshun pekos nezod lekiv rosun nemuz resod sibiv ligun kivuz
	Trained-item list		
	A	B	C
	Diversity: 0.133	Diversity: 0.269	Diversity: 0.125
Medium frequency inflection (12 words) Suffix: "esh"	Words containing "consistent rime cues" Predictability: 0.8 3 families of 4 words each: napod paniv koshun nezod tepiv rosun resod lekiv ligun moshod sibiv batun	Words containing "consistent rime cues" Predictability: 1.0 3 families of 2 words each: napod tepiv rosun nezod sibiv ligun Words containing "inconsistent rime cues" • 3 words rhyming with words that take the high frequency inflection (gomoz, gagul, zesig). Predictability: 0.11. • 3 words rhyming with words taking low frequency inflections (kunus, gomil pakom). Predictability: 0.5	Words containing "consistent rime cues" Predictability: 1.0 3 families of 4 words each: timut darok shashib zumut kupok nilib gokut gudjok gofib sozut rofok mutib
Low frequency inflections (3 inflections × 4 words each) Suffixes: "ev," "ak," "ur"	Diversity (of each suffix): 0.194 None of the words rhyme with each other: meshus shibil zufom Predictability: for nine words- 1.0; for three words (meshus, shibil, zufom)- 0.5	dipem tegas lidek mikal getav nised	Diversity (of each suffix): 0.167 sapor nerud rinit Predictability: 1.0

Table 2. Number of untrained words in each test, for each group (A, B, C).

	A	B	C
Words with unambiguous rime cues	49	49	73
Words with ambiguous rime cues	24	24	0
Words without rime cues	23	23	23
Total	96	96	96

Procedure

Presentation of stimuli and recording of vocal responses was done using E-Prime stimulus presentation software (v.2.0, Psychological Software Tools, Inc.). Recording of vocal responses was triggered using the "voice key" feature on the Serial Response Box (Psychological

Software Tools, Inc.). Training took place over the course of 6 sessions. The mean interval between consecutive sessions was 2.83 days (range: 1–5). In 7 unavoidable cases (2.8% of sessions) there was a longer interval of up to 7 days.

The first session began with an instruction block (see Figure 1), which exposed participants to the trained words in both their singular and plural forms, together with their meanings. In the instruction block each of the 48 training items was presented once, in both its singular and plural forms (for the design of each trial in the instruction block, see Figure 2(a)). Each trial began with the presentation of a fixation cross. When the participant pressed the space bar the singular form of a word was presented aurally together with an image of a real object (e.g. an apple) on the computer's screen. Participants were informed at the beginning of this block that these were the objects the items referred to. The images were included in order to increase the resemblance of the artificial language to natural languages. The use of familiar objects (which correspond to words in participants' L1) makes this artificial language paradigm more similar to learning a second language. The singular form was followed by a visual cue consisting of two asterisks (**) in the centre of the screen for one second, indicating that the plural form of the word would soon be presented. The plural form was then presented aurally, followed by the presentation of a question mark in the centre of the screen, indicating that participants were to repeat the plural form they had just heard. This cue remained on the screen for a maximal duration of four seconds, or until a vocal response was detected. See Figure 2(a) for the design of this task.

Trained-item tests that required judgments of correctly and incorrectly inflected plural forms were presented both before and after each training session (see Figure 2). The design of trials in this task is presented in Figure 1(b). Each of the 48 trained words was tested once. In each trial a visual cue was presented for 500 ms, followed by an aural presentation of the singular form of a trained stem. No images were presented during this task. This was followed by a visual cue consisting of two asterisks (**) for one second, and then a plural form of the same word presented aurally. Half of the 48 plural forms presented were correct, and half

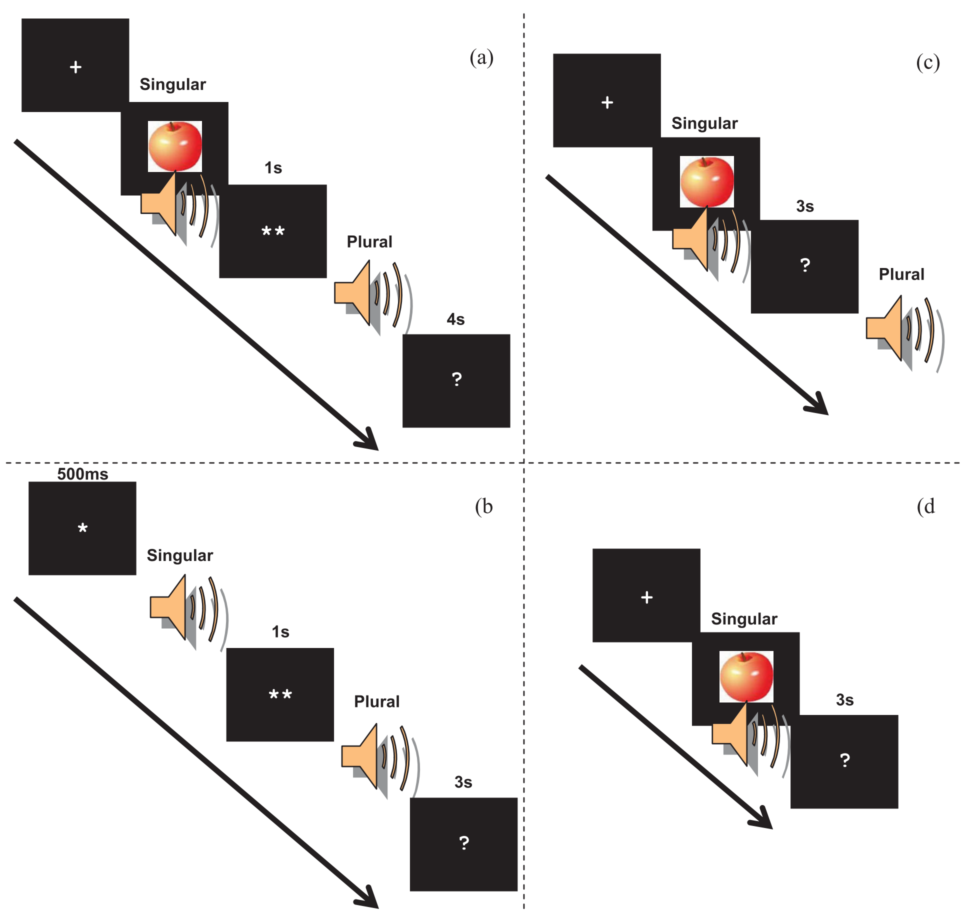

Figure 1. Design of trials in (a) instruction block, (b) trained-items tests, (c) training blocks, (d) untrained items test.

were incorrect. Incorrect inflections were created by adding one of the other suffixes, presented during training, to the singular form. For the two tests within each session, each word was presented once with a correct inflection and once with an incorrect inflection. Presentation of correct and incorrect inflections was counterbalanced across participants. Presentation of the plural form was followed by the presentation of a question mark in the centre of the screen, indicating that a response was required. Participants were instructed to press "1" if the plural form was correct, and to press "2" if not. The maximal reaction time allowed was 3 s.

In each of the six sessions, participants also completed training blocks (Figure 2), in which they practiced the production of inflected forms on the 48 trained words. During training each trial began with the presentation of a fixation cross followed by aural presentation of the singular form of a trained word and a visual presentation of the object the trained word referred to. This was followed by a question mark, prompting participants to pronounce the plural form of the word they had just heard. Once recording of the response terminated, or three seconds elapsed without a vocal response, the correct plural form of the word was presented aurally, as feedback regardless of the accuracy of the response. This was done to ensure an equal amount of exposure to the correct inflections among all participants. See Figure 1(c) for the design of this task. The training session consisted of five blocks, in each of which each trained word was presented once.

In the first and last sessions (sessions 1 and 6), after the second trained-items test, participants were also tested on producing the inflections of untrained words. In these tests participants heard untrained words in their singular form, and were asked to produce their

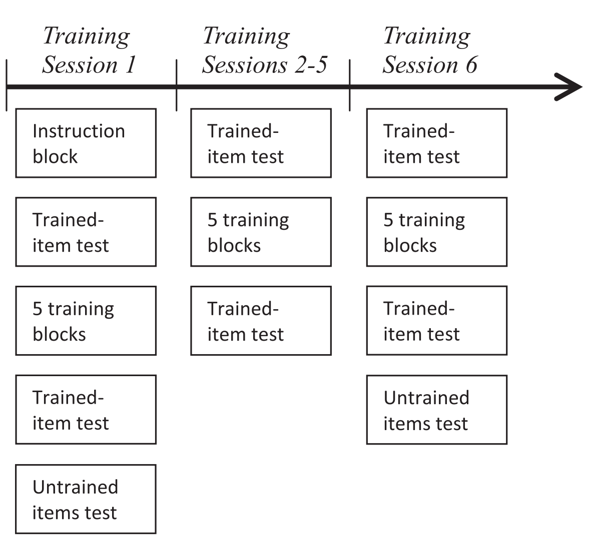

Figure 2. Overall design of the experiment.

plural inflection. Trials in this task resembled trials in the training blocks with the exception that participants did not receive any feedback (see Figure 1(d)).

At the end of the 6th session participants received a written questionnaire to assess their explicit knowledge of the inflections' frequencies and the phonological cues embedded in trained words. Participants were asked which suffixes they remembered, what their estimates of the suffixes' relative frequencies were, what information they had been using to inflect untrained words, and whether they had noticed any regularities in the words they had trained on.

Statistical analyses

Analyses were performed using IBM SPSS Statistics software (v.17 and 19).

Trained-item tests

Participants' accuracy and reaction times during trained-item tests were examined separately using repeated-measures ANOVAs, in order to characterise their learning of the trained inflections and to determine how this learning was affected by suffixes' frequencies. As mentioned above, lists A and B had been designed to differ in terms of affix diversities, whereas list C had been designed to be similar to list A in terms of distributions of diversities, and differ on affix predictability. We therefore performed separate comparisons for groups A and B, to examine effects of diversity, and for groups A and C, to examine effects of predictability. All comparisons were performed both by-participant, and by-item, resulting in a total of 8 ANOVAs. Within-subjects factors in all ANOVAs (A & B, accuracy; A & B, RT; A & C, accuracy; A & C, RT) were session (1–6) and "test number" (before or after training). Suffix type frequency served as a within-subject factor in by-participant analyses, and as a "between-subjects" factor in by-item analyses, since each word received only one suffix. In all ANOVAs the Greenhouse-Geisser correction was applied for sphericity values lower than 0.75, and the Huynh-Feldt correction was applied for sphericity values greater than 0.75 (see Field, 2005).

Since words containing "inconsistent rime cues" had been included in lists A and B only in order to reduce affix predictability in these lists and to enable the examination of the differences between the effects of "perfect" (i.e, deterministic) and imperfect cues, and were not of interest in themselves, these words were excluded from analyses. However, prior to their exclusion we compared performance (both accuracy and reaction time) on words containing inconsistent rime cues with performance on words containing consistent rime cues receiving the same suffixes: i.e. within the high -frequency suffix for group A, and the medium-frequency-suffix for group B. See Results for the outcome of this comparison.

Inflections of untrained words

Data collected from one participant in group C were omitted from analyses of untrained words because this participant deliberately used novel suffixes not presented during training, resulting in data that were incomparable to those obtained from other participants.

Distributions of inflections: In order to examine experience-dependent changes in participants' application of affixes to untrained words, responses were classified by suffix type frequency category (High, Medium, Low). Proportions of responses using suffixes in each frequency category were calculated for each participant in each session, and entered as the dependent variable in repeated-measures ANOVAs, performed separately for words containing rime cues, and for words without rime cues. Within-subject factors were session (1st, 6th) and affix-frequency (High, Medium, Low). Group (A, B, and C) was the between-subject factor. Additionally, to perform by-item analyses, proportions of responses using suffixes in each frequency category were calculated for each untrained word, and examined separately for words containing rime cues and for words without rime cues, again using repeated-measures ANOVAs. It should be noted that in these analyses session was a "between-subjects" factor, since each untrained word was presented only once, either in the 1st session, or in the 6th session.

Effects of suffix predictability- analysis of "optimal responses": In order to examine participants' sensitivity to phonological cues and their reliance on predictability we analysed their performance when inflecting untrained words containing rime cues. This was done by examining the extent to which participants inflected these words using suffixes with the highest predictability, based on the phonological cues. Inflections using the suffix with the highest predictability are referred to as "optimal" responses. Proportions of "optimal" responses (i.e. the number of optimal responses divided by the number of words containing rime cues, which was 73; see Table 2) were calculated for each participant in each session. These values served as the dependent measure in two repeated-measures ANOVAs, one examining effects by-participant, and the other examining effects by-item. Session (1st, 6th) and group (A, B, C) were the independent variables in both of these analyses. Session was entered as a within-subject factor in the by-participant analysis, and as a "between-subjects" factor in the by-item analysis, since each untrained word was presented only once, either in the 1st session, or in the 6th session.

Additionally, the proportions of optimal responses thus calculated were compared with "baseline" proportions of optimal responses, that is, proportions of optimal responses that might be expected if participants were ignoring the cues embedded in the untrained words, and instead relying on suffixes' frequencies or diversities. These "baseline" proportions were obtained by multiplying for each suffix the proportion of untrained words for which it serves as the optimal response by that suffix's frequency or diversity, and summing the resulting proportions across suffixes. For example, if the optimal response for one third of the untrained words involves using the high-frequency suffix (which is applied to 0.5 of trained words) then the baseline probability of applying this suffix is 0.33×0.5. Similarly, if the optimal response of another third of the untrained words involves using the medium-frequency suffix (which was applied to 0.25 of trained words), then the baseline probability of applying this suffix is: 0.33×0.25. These probabilities are then summed across suffixes. See Table 5.

Comparison of reliance on suffix type frequency and diversity -"cosine similarities": In order to compare between participants' reliance on suffix type frequency and their reliance on suffixes' diversity, for inflecting untrained words without rime cues, we used a measure of "cosine similarity." This assesses the similarity between the distribution of responses participants had provided, and the distributions which could be expected if participants relied solely on either frequency or diversity. The term "cosine similarity," which is used mainly in the field of Information Retrieval (e.g. Salton & Lesk, 1968), refers to the cosine of the angle between two non-zero vectors. This is given by: $\cos\theta = a \cdot b / \parallel a \parallel \parallel b \parallel$ where $\cdot$ indicates the inner product (i.e. $a \cdot b = \sum a_i b_i$), and $||x||$ indicates the magnitude of a vector x, equivalent to $\sqrt{x \cdot x}$. The value of the cosine ranges from −1, indicating that the two vectors are exactly opposite to each other, to 1, indicating that the two vectors are perfectly aligned, with 0 indicating orthogonality. Cosines between distributions of responses and expected distributions based on suffix type frequency and suffix diversity were calculated for each participant in each session (1st or 6th), for words without rime cues, which had been presented for this purpose. The calculated values were entered into an ANOVA, in which session (1st, 6th) and "affix property" (suffix type frequency vs. suffix diversity) served as within-subject factors, and group as the between-subject factor.

Questionnaires

Participants' estimates of suffixes' frequencies, as provided in their responses to the questionnaires presented at the end of the 6th session were compared to actual frequencies using Pearsons' chi-squared test for goodness of fit. As for knowledge of phonological cues, proportions of "optimal responses" were compared between participants who reported noticing regularities in words' rimes and participants who did not, using *t*-tests for independent-samples.

Results

Trained-item tests

Participants' judgements concerning inflections of trained words containing consistent rime cues were analysed with repeated-measures ANOVAS, with session (1–6), "test number" (when performance was assessed; before training = "test 1", after training = "test 2"), and suffix type frequency (High, Medium or Low) as within subject variables, and group as the between-subjects variable. Separate ANOVAs were performed for accuracy (Figure 3) and reaction times (Figure 4). Performance in groups A and B was compared to examine the effect of diversity, and groups A and C were compared separately to examine effects of predictability. All comparisons were performed both by-participant and by-item, resulting in a total of 8 ANOVAs. Only effects which were significant both in by-participant and by-item analyses are reported.

Prior to the analysis of words containing consistent rime cues, a preliminary analysis was conducted comparing these words to words containing inconsistent rime cues. These comparisons yielded significant differences between the two types of words in accuracy (main effect: $F_{(1,33)} = 15.01$, $p < 0.001$, $\eta_p^2 = 0.31$) and reaction times (main effect: $F_{(1,25)} = 5.02$, $p < 0.05$, $\eta_p^2 = 0.17$) across the high- and medium-frequency suffixes. These differences, suggesting an effect of predictability on learning of trained inflections, supported the exclusion of words containing inconsistent rime cues from all other analyses of trained items.

Learning of trained items: improvements over and within sessions

Main effects of session were found in all comparisons (see Table 3), indicating that participants' performance

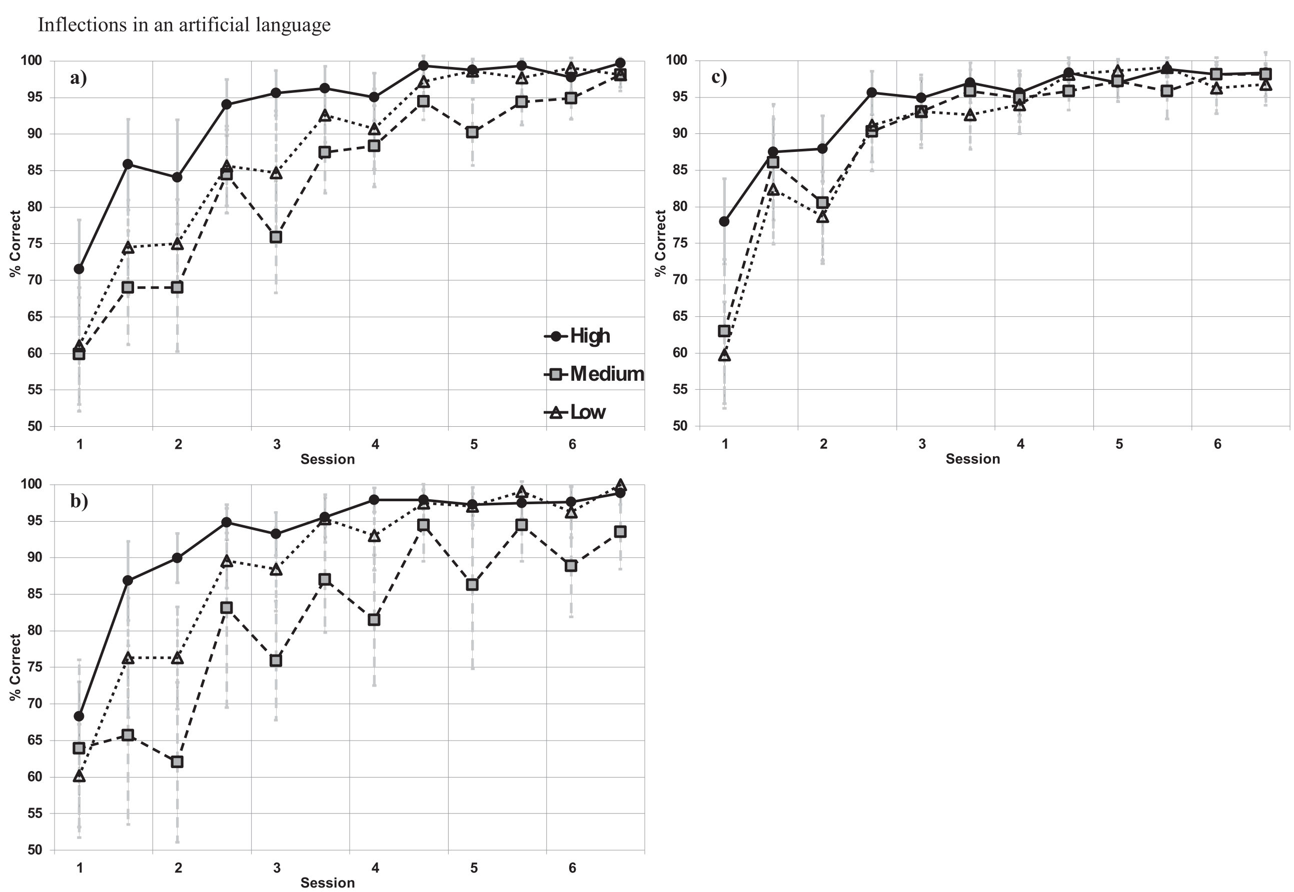

Figure 3. Accuracy of responses on trained items. (a) group A, (b) group B, (c) group C. Horizontal axis represents sessions (two tests per session), and vertical axis represents percent of correct responses. Error bars indicate limits of 95% confidence intervals.

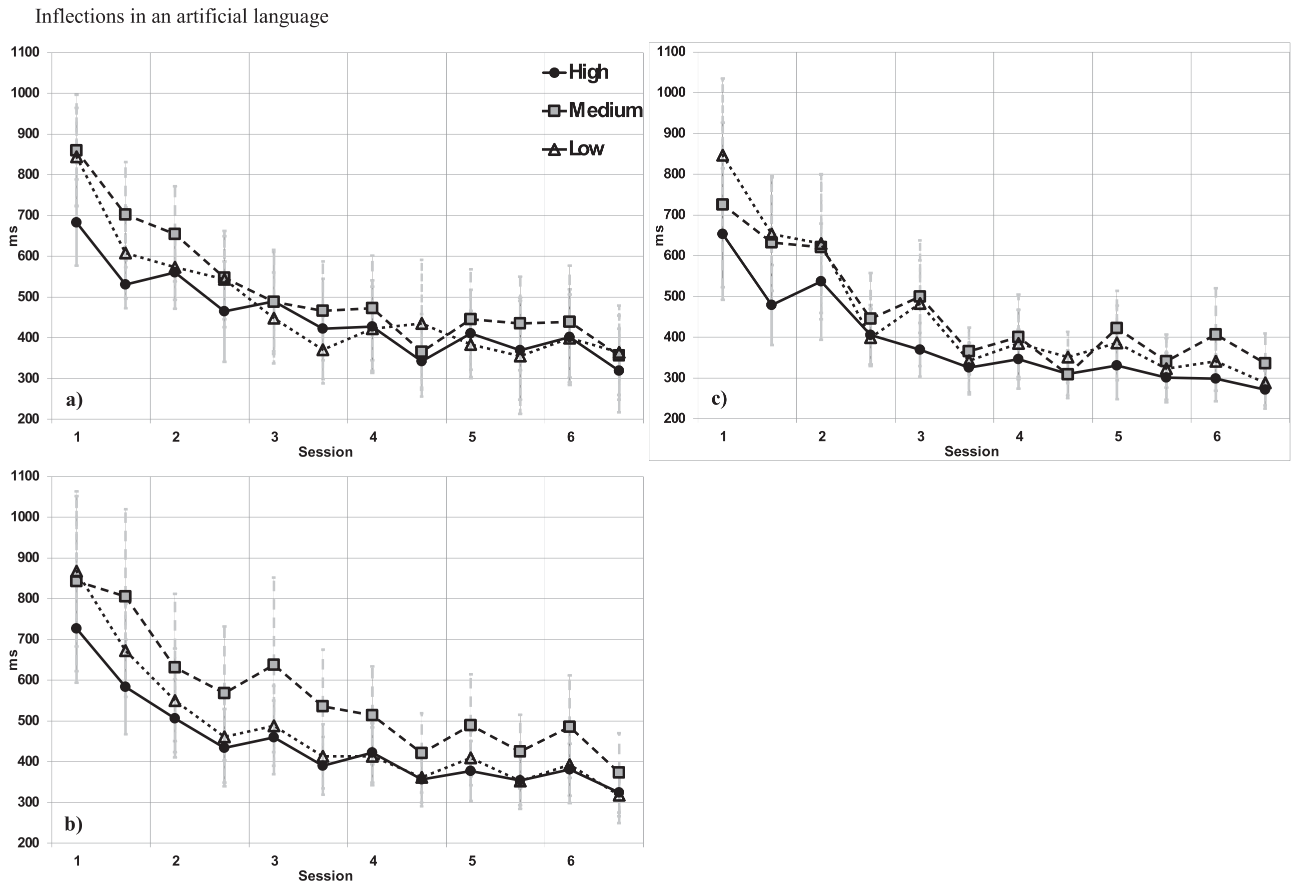

Figure 4. Reaction times on trained item. (a) group A, (b) group B, (c) group C. Horizontal axis represents sessions (two tests per session), and vertical axis represents reaction time in ms. Error bars indicate limits of 95% confidence intervals.

improved over the course of training both for accuracy and reaction time (see Figures 3 and 4). Participants' performance also improved from the beginning to the end of each session, resulting in a significant main effect of test in all comparisons (see Table 3). The session × test interaction was also significant in all analyses, due to greater improvement in early as compared to later sessions (see Figures 3 and 4). (see Table 3).

Learning of trained items: effects of suffix type frequency

All analyses of trained items showed main effects of suffix type frequency, as well as session × frequency interactions (see Table 3). Pairwise comparisons were performed to further explore these effects. These comparisons revealed higher accuracy on high-frequency inflections compared to medium- and low-frequency inflections in groups A and B. On the other hand, reaction times for low-frequency inflections were shorter in these groups than reaction times for medium frequency inflections, and did not differ from reaction times for high-frequency inflections. Additionally, group B differed from group A in that accuracy on low-frequency inflections was higher than for medium-frequency inflections in the former, but not in the latter. None of the comparisons yielded significant results in group C. See Table 4.

Learning of trained items: effects of experimental group

A main effect of experimental group on accuracy was found only for the comparison between groups A and C, with participants in group C responding more accurately overall than participants in group A. Additionally, the session × group interaction on accuracy was also significant in the analysis of groups A and C (see Table 3), due to the fact that differences in accuracy were no longer evident by the 5th session in group A (see Figure 3(a,c)).

Learning of trained items: summary

Altogether, the results of the trained-item tests show higher accuracy on words receiving high frequency suffixes, and higher accuracy and faster improvement in group C, which was trained on inflections that were completely predictable. These results suggest that both suffix type frequency and predictability affect the learning of trained items.

Table 3. Effects found in analyses of trained-item tests. Only comparisons that yielded significant results both in by-participant and by-item analyses are presented.

Effect	Comparison	F	df_{effect}	df_{error}	p (uncorrected)	η_p^2Z
Session	A vs. B, Acc, by-Participant	257.31	3.37	111.06	$<10^{-6}$	0.89
	A vs. B, RT, by-Participant	50.97	2.63	86.66	$<10^{-6}$	0.61
	A vs. C, Acc, by-Participant	262.78	3.27	111.06	$<10^{-6}$	0.89
	A vs. C, RT, by-Participant	50.53	2.41	81.79	$<10^{-6}$	0.6
	A vs. B, Acc, by-Item	191.23	3.55	319.18	$<10^{-6}$	0.68
	A vs. B, RT, by-Item	294.44	4.17	375.7	$<10^{-6}$	0.77
	A vs. C, Acc, by-Item	191.77	2.98	268.53	$<10^{-6}$	0.68
	A vs. C, RT, by-Item	305.93	4.07	366.54	$<10^{-6}$	0.77
Test	A vs. B, Acc, by-Participant	200.90	1	33	$<10^{-6}$	0.86
	A vs. B, RT, by-Participant	56.72	1	33	$<10^{-6}$	0.63
	A vs. C, Acc, by-Participant	155.14	1	34	$<10^{-6}$	0.82
	A vs. C, RT, by-Participant	48.89	1	34	$<10^{-6}$	0.59
	A vs. B, Acc, by-Item	271.82	1	90	$<10^{-6}$	0.75
	A vs. B, RT, by-Item	171.62	1	90	$<10^{-6}$	0.66
	A vs. C, Acc, by-Item	187.70	1	90	$<10^{-6}$	0.68
	A vs. C, RT, by-Item	197.00	1	90	$<10^{-6}$	0.69
Session × Test	A vs. B, Acc, by-Participant	11.19	3.04	100.32	$<10^{-5}$	0.25
	A vs. B, RT, by-Participant	3.19	3.36	111.03	0.022	0.09
	A vs. C, Acc, by-Participant	20.6	2.60	88.37	$<10^{-6}$	0.38
	A vs. C, RT, by-Participant	3.70	2.50	84.94	0.02	0.10
	A vs. B, Acc, by-Item	20.28	3.40	306.22	$<10^{-6}$	0.18
	A vs. B, RT, by-Item	10.86	3.95	355.04	$<10^{-6}$	0.11
	A vs. C, Acc, by-Item	36.40	3.27	294.23	$<10^{-6}$	0.29
	A vs. C, RT, by-Item	10.35	3.73	335.52	$<10^{-6}$	0.10
Frequency	A vs. B, Acc, by-Participant	62.86	1.75	57.88	$<10^{-6}$	0.66
	A vs. B, RT, by-Participant	26.38	2	66	$<10^{-6}$	0.44
	A vs. C, Acc, by-Participant	26.52	2	67.86	$<10^{-6}$	0.44
	A vs. C, RT, by-Participant	26.36	2	68	$<10^{-6}$	0.44
	A vs. B, Acc, by-Item	35.55	2	90	$<10^{-6}$	0.44
	A vs. B, RT, by-Item	22.66	2	90	$<10^{-6}$	0.34
	A vs. C, Acc, by-Item	16.78	2	90	$<10^{-5}$	0.27
	A vs. C, RT, by-Item	14.48	2	90	$<10^{-5}$	0.24
Session × Frequency	A vs. B, Acc, by-Participant	7.84	4.95	163.45	$<10^{-5}$	0.19
	A vs. B, RT, by-Participant	3.10	4.48	147.88	0.014	0.09
	A vs. C, Acc, by-Participant	7.41	5.15	175.03	$<10^{-5}$	0.18
	A vs. C, RT, by-Participant	4.76	5.96	202.64	$<10^{-3}$	0.12
	A vs. B, Acc, by-Item	5.06	7.09	319.18	$<10^{-4}$	0.10
	A vs. B, RT, by-Item	3.37	8.35	375.70	$<10^{-3}$	0.07
	A vs. C, Acc, by-Item	5.16	5.97	268.53	$<10^{-4}$	0.10
	A vs. C, RT, by-Item	4.43	8.15	366.54	$<10^{-4}$	0.09
Group	A vs. C, Acc, by-Participant	6.94	1	34	0.013	0.17
	A vs. C, Acc, by-Item	16.45	1	90	$<10^{-3}$	0.16
Session × Group	A vs. C, Acc, by-Participant	4.98	3.27	111.06	0.002	0.13
	A vs. C, Acc, by-Item	3.9	2.98	268.53	0.01	0.04
Session × Test × Frequency	A vs. B, Acc, by-Participant	2.55	6.14	202.61	0.02	0.07
	A vs. B, Acc, by-Item	3.53	6.81	306.22	0.001	0.07

Inflections of untrained words

Distributions of inflections

Distributions of inflections of untrained words were examined using separate repeated-measures ANOVAs for untrained words containing rime cues and untrained words without rime cues. In these by-participant and by-item analyses the independent factors were suffix type frequency (High, Medium, or Low), session (1st or 6th), and group.

For untrained words containing rime cues the main effect of suffix type frequency was significant (by-participant: $F_{(2,100)} = 132.60$, $p < 0.001$, $\eta_p^2 = 0.73$; by-item: $F_{(1.82,784.48)} = 192.88$, $p < 0.001$, $\eta_p^2 = 0.31$), indicating that the high-frequency suffix was applied most often, followed by the medium-frequency suffix, which was applied more often in turn than the low-frequency suffixes combined. The suffix type frequency × session interaction was not significant (by-participant: $F_{(2,100)} = 1.13$, N.S.; by-item: $F_{(1.82,784.48)} = 0.82$, N.S.); however, the suffix type frequency × session × group interaction was significant (by-participant: $F_{(4,100)} = 4.69$, $p < 0.01$, $\eta_p^2 = 0.16$; by-item: $F_{(3.63,784.48)} = 3.35$, $p < 0.05$, $\eta_p^2 = 0.02$). Following up on this interaction, separate ANOVAs for each of the three groups yielded a significant suffix type frequency × session interaction only for group B (by-participant: $F_{(2,34)} = 6.80$, $p < 0.01$, $\eta_p^2 = 0.29$; by-item: $F_{(1.84,264.59)} = 7.55$, $p < 0.01$, $\eta_p^2 = 0.05$). Results of pairwise comparisons indicated that this was due to an increase in the application of the medium-frequency suffix from the 1st to the 6th session (by-participant: $t_{(17)} = 3.76$, $p < 0.01$;

Table 4. Results of pairwise comparisons examining effects of affix type frequency on learning of trained inflections.

Group	Comparison	t	d.f.	p (2-tailed, uncorrected)
A	High vs. Medium, Accuracy, by Participant	7.05	17	$<10^{-5}$
	High vs. Low, Accuracy, by Participant	4.22	21,517	0.001
	High vs. Medium, RT, by Participant	5.29	17	$<10^{-4}$
	Low vs. Medium, RT, by Participant	2.47	17	0.024
	High vs. Medium, Accuracy, by Item	4.94	15.43	$<10^{-3}$
	High vs. Low, Accuracy, by Item	2.41	14.39	0.03
	High vs. Medium, RT, by Item	3.47	17.05	0.003
	Low vs. Medium, RT, by Item	1.71	22	0.10
B	High vs, Medium, Accuracy, by Participant	7.16	17	$<10^{-5}$
	High vs. Low, Accuracy, by Participant	4.34	17	$<10^{-3}$
	Low vs. Medium, Accuracy, by Participant	5.66	17	$<10^{-4}$
	High vs. Medium, RT, by Participant	4.51	17	$<10^{-3}$
	Low vs. Medium, RT, by Participant	4.62	17	$<10^{-3}$
	High vs. Medium, Accuracy, by Item	4.11	9.29	0.002
	High vs. Low, Accuracy, by Item	2.43	34	0.021
	Low vs. Medium, Accuracy, by Item	2.77	11.23	0.018
	High vs. Medium, RT, by Item	4.65	31	$<10^{-5}$
	Low vs. Medium, RT, by Item	3.26	19	0.004

Note: Only comparisons that yielded significant results both in by-participant and by-item analyses are presented.

by-item: $t_{(109.67)} = 4.37$, $p < 0.001$), as can be seen in Figure 5(a).

The analysis of untrained words without rime cues also showed a significant main effect of suffix type frequency (by-participant: $F_{(1.73,86.41)} = 28.17$, $p < 0.001$, $\eta_p^2 = 0.36$; by-item: $F_{(2.00,263.70)} = 26.87$, $p < 0.001$, $\eta_p^2 = 0.17$). However, here the suffix type frequency × session interaction was significant (by-participant: $F_{(2,100)} = 34.60$, $p < 0.001$, $\eta_p^2 = 0.41$; by-item: $F_{(2.00,263.70)} = 11.67$, $p < 0.001$, $\eta_p^2 = 0.08$), while the three-way interaction was not (by-participant: $F_{(4,100)} = 1.88$, N.S.; by-item: $F_{(4.00,263.70)} = 0.68$, N.S.). Results of follow-up pairwise comparisons showed that the interaction was due to an increase in the application of low-frequency suffixes from the first session to the sixth session (by-participant: $t_{(52)} = 9.53$, $p < 0.001$; by-item: $t_{(100.12)} = 5.78$, $p < 0.001$), accompanied by a decrease in the application of the high-frequency suffix (by-participant: $t_{(52)} = 3{,}72$, $p < 0.001$; by-item: $t_{(129.60)} = 2.22$, $p < 0.05$), with no change in the application of the medium-frequency suffix (by-participant: $t(52) = 0.44$, N.S.; by-item: $t_{(125.04)} = 0.29$, N.S.). As a result, whereas in the first session the high-frequency suffix was applied more often than the low-frequency suffixes (by-participant: $t_{(52)} = 6.22$, $p < 0.001$; by-item: $t_{(68)} = 7.16$, $p < 0.001$), which did not differ from the medium-frequency suffix (by-participant: $t_{(52)} = 0.41$, N.S.; by-item: $t_{(68)} = 0.59$, N.S.), by the sixth session the low-frequency suffixes were applied as often as the high-frequency suffix (by-participant: $t_{(52)} = 0.80$, N.S.; by-item: $t_{(68)} = 0.57$, N.S.), and more often than the medium-frequency suffix (by-participant: $t_{(52)} = 6.21$, $p < 0.001$; by-item: $t_{(68)} = 4.28$, $p < 0.001$).

To summarise, these analyses show an initial tendency to apply the high frequency suffix to untrained words with and without rime cues. For words without rime cues this tendency decreased over time, together with an increase in the application of the low frequency suffix, resulting in similar application of both high frequency and low frequency suffixes in the 6th session. See Figure 5(b).

Effects of suffix predictability: analysis of "optimal responses"

To examine effects of suffix predictability, we probed participants' reliance on rime cues in the inflection of untrained words that contain rime cues. Specifically, we analysed the "optimal responses" for these items (for a detailed description of optimal responses, see Methods section). The proportion of optimal responses provided by each participant in each session was entered into a repeated-measures ANOVA for the by-participant analysis, in which session (1st, 6th) was the within-subject factor, and group (A, B, C) was the between-subjects factor. Additionally, the proportion of optimal responses provided for each word was entered into an ANOVA for the by-item analysis, in which both session and group served as "between-subjects" factors.

These analyses yielded main effects for both factors, that is, session and group. The main effect of session (by-participant: $F_{(1,50)} = 97.47$, $p < 0.001$, $\eta_p^2 = 0.66$; by-item: $F_{(1,432)} = 35.12$, $p < 0.001$, $\eta_p^2 = 0.08$) was due to an increase in optimal responses from the 1st to the 6th session. There was also a main effect of group (by-participant: $F_{(2,48)} = 12.61$, $p < 0.001$, $\eta_p^2 = 0.34$; by-item: $F_{(2,432)} = 20.85$, $p < 0.001$, $\eta_p^2 = 0.09$). Pairwise comparisons indicated that participants in group C provided significantly more optimal responses than participants in group B (by-participant: $p < 0.05$; by-item: $p < 0.001$. See Figure 6. Proportions of optimal responses exceeded the baseline proportions (expected application of suffix based on frequency or diversity alone) by the end of the 1st session in group C, but not in groups A and B. By the end of the 6th session proportions of optimal responses exceeded baseline levels in all groups.

Overall, the analysis of optimal responses shows reliance on suffix predictability by the end of the first session in group C, which had more predictable cues,

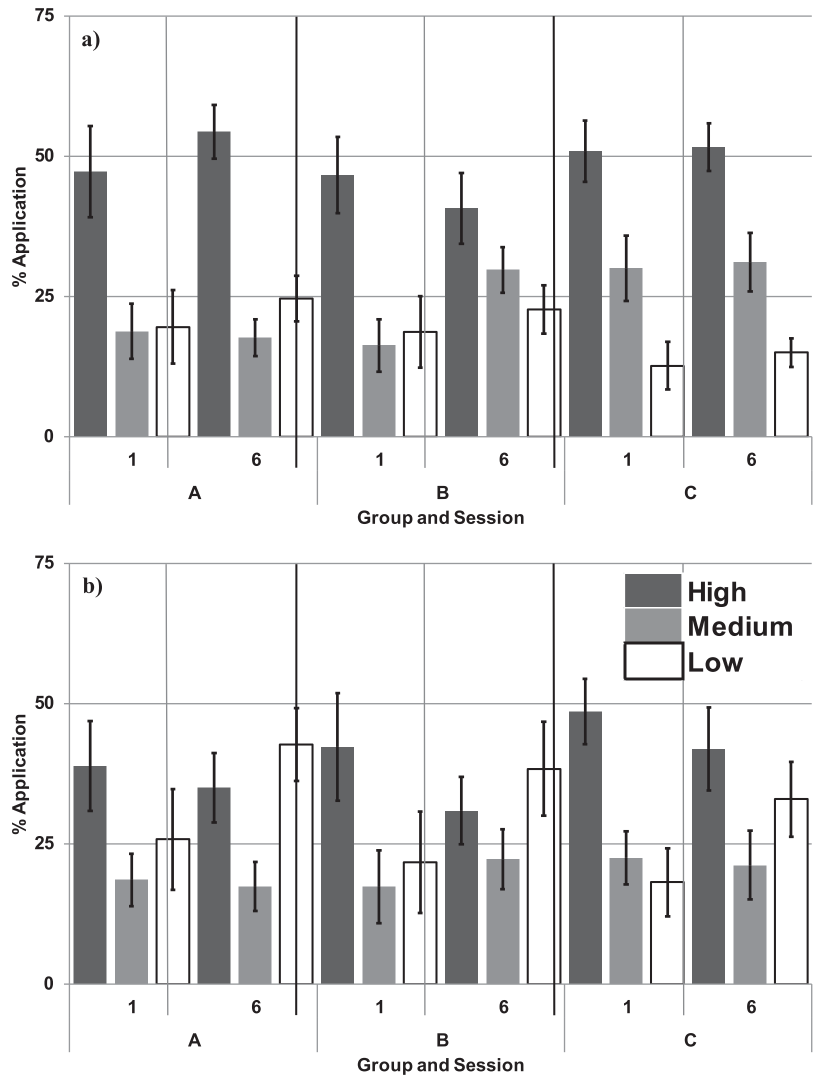

Figure 5. Distribution of inflections of untrained words containing rime cues (a) and without rime cues (b). Percent of untrained words inflected using High- Medium- and Low-frequency suffixes. Horizontal lines mark expected use based on the suffixes' frequency among trained items: 50% for the high frequency suffix, 25% for the medium frequency suffix and for the sum of low frequency suffixes. Error bars indicate the limits of a 95% confidence interval about the mean.

with an increase in reliance on phonological cues during training in all groups. See Table 5.

Comparison of reliance on suffix type frequency and diversity: "cosine similarities"

In order to compare participants' reliance on suffix type frequency with their reliance on affixal diversity for the inflection of untrained words, we examined "cosine similarities" between actual distributions of responses and the expected distributions based on suffix type frequency and diversity, for words without rime cues. These cosine similarity measures were the dependent variable in a repeated-measures ANOVA, with session and "affix property" (suffix type frequency vs. suffix diversity) as within-subject factors, and group as a between-subject factor.

The main effect of "affix property" was significant ($F_{(1,50)} = 17.70$, $p < 0.001$, $\eta_p^2 = 0.26$), due to a greater reliance on suffix type frequency across groups and sessions. However, the session × "affix property" interaction was also significant ($F_{(1,50)} = 50.83$, $p < 0.001$, $\eta_p^2 = 0.50$), as was the three-way interaction ($F_{(2,50)} = 4.20$, $p < 0.05$, $\eta_p^2 = 0.14$). Consequently, two sets of analyses were performed: separate analyses for suffix type frequency and for suffixal diversity, with group and session as the independent variables, and separate analyses for each session, with group and "affix property" as the independent variables.

In the first set of analyses the main effect of group on reliance on affix type frequency was found to be marginally significant ($F_{(2,50)} = 3.19$, $p = 0.05$, $\eta_p^2 = 0.11$), with no effects of session, indicating that there was no change

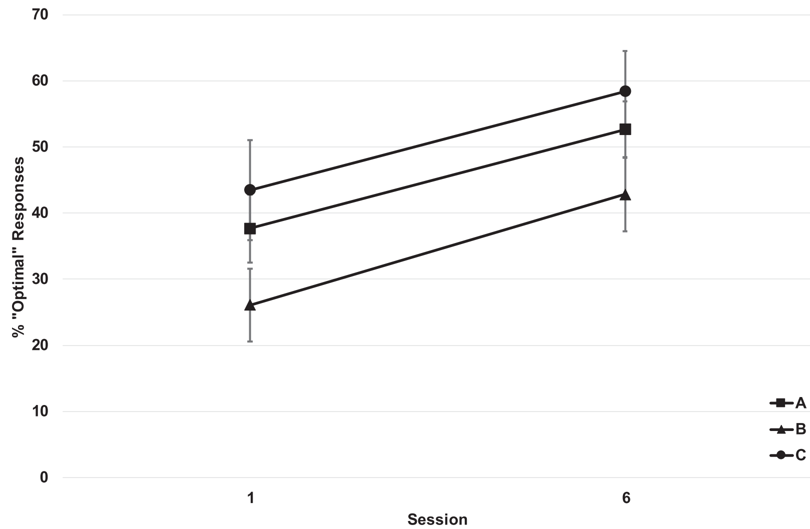

Figure 6. Percent of "optimal" responses in untrained words with rime cues, presented by group (A, B, C) and session (1st, 6th). Error bars indicate the limits of a 95% confidence interval about the mean.

Table 5. Results of comparisons between actual proportions of "optimal" inflections of untrained words and the expected proportions based on suffix type frequency or diversity.

Group	Session	Proportion of "optimal" responses	Expected proportion based on		*t*	d.f.	*p* (2-tailed, uncorrected)
			Frequency	Diversity			
A	1	0.38	**0.33**	0.26	1.92	17	0.071
	6	0.53			9.75	17	$<10^{-6}$
B	1	0.26	0.25	**0.26**	0.22	17	0.83
	6	0.43			6.52	17	$<10^{-5}$
C	1	0.44	**0.28**	0.22	4.48	16	$<10^{-3}$
	6	0.58			10.72	16	$<10^{-6}$

Note: Proportions of optimal responses were compared to the higher of these expected proportions (indicated by bold font).

in the reliance on suffix type frequency. In contrast, reliance on suffixal diversity differed between groups and sessions (main effect of group: $F_{(2,50)} = 19.46$, $p < 0.001$, $\eta_p^2 = 0.44$; main effect of session: $F_{(1,50)} = 60.68$, $p < 0.001$, $\eta_p^2 = 0.55$; interaction: $F_{(2,50)} = 6.64$, $p < 0.01$, $\eta_p^2 = 0.21$). Results of pairwise comparisons indicated an increase in reliance on suffix diversity from the 1st session to the 6th session in groups A and B, but not in group C (A: $t_{(17)} = 6.49$, $p < 0.001$; B: $t_{(17)} = 5.71$, $p < 0.001$; C: $t_{(16)} = 1.59$, N.S.).

Separate analyses for the 1st and 6th session yielded a main effect of "affix property" in the 1st session ($F_{(1,50)} = 52.68$, $p < 0.001$, $\eta_p^2 = 0.51$), due to greater reliance on suffix type frequency. The main effect of group was also significant ($F_{(2,50)} = 9.02$, $p < 0.001$, $\eta_p^2 = 0.27$), as was the interaction of group and "affix property" ($F_{(2,50)} = 5.13$, $p < 0.01$, $\eta_p^2 = 0.17$). In contrast, no effects of "affix property" were found in the 6th session, and only the main effect of group was significant ($F_{(2,50)} = 5.51$, $p < 0.01$, $\eta_p^2 = 0.18$).

Altogether, the analysis of cosine similarities indicates that in the 1st session there was more reliance on affix type frequency than on suffix diversity, for untrained words without rime cues. However, reliance on suffix diversity increased from the 1st session to the 6th session ($t_{(52)} = 7.15$, $p < 0.001$), especially in groups A and B, so that by the 6th session there was similar reliance on affix type frequency and suffix diversity. See Figure 7.

Questionnaires

In order to test whether participants developed explicit awareness of suffix type frequencies and of the phonological cues embedded in trained words, we analysed the results of the questionnaires administered at the end of the 6th session. By the end of the 6th session participants were able to provide good explicit estimates of suffix type frequencies. This is indicated by the fit between participants' frequency estimates and the actual frequencies of inflections of trained stimuli. Pearson's chi-squared tests for goodness of fit showed no significant differences between the two ($\chi^2_{(4)} = 1.14$, $p = 0.89$; group A: $\chi^2_{(4)} = 2.15$, $p = 0.71$; group B: $\chi^2_{(4)} = 1.18$, $p = 0.88$; group

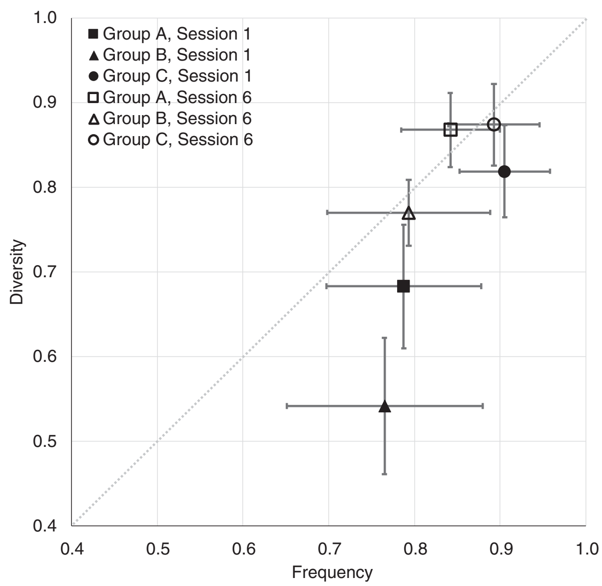

Figure 7. Cosine similarities for reliance on affix diversity and affix type frequency in untrained words without rime cues, presented by group (A, B, C) and session (1st, 6th). Error bars indicate the limits of a 95% confidence interval about the mean. The diagonal represents similar reliance on diversity and type frequency.

C: $\chi^2_{(4)} = 3.68, p = 0.45$). However, most participants (17/18 in each group) reported they did not rely on knowledge of frequency distributions during the inflection of untrained words.

Only a minority of participants (24%) mentioned noticing regularities concerning words' rimes. In order to determine whether these participants were relying more heavily on knowledge of phonological cues than participants who had not reported noticing regularities in words' rimes, proportions of optimal responses (see above) provided by these two groups of participants were compared. Results of these comparisons indicated greater reliance on phonological cues among participants who had noticed regularities in words' rimes. This was true for the 6th session but not for the 1st (1st session- $t_{(51)} = 1.03$, N.S; 6th session- $t_{(35.36)} = 3.58$, $p < 0.01$).

Summary and discussion

This study examined the effects of suffix type frequency, suffix phonological predictability, and suffix diversity on the learning and generalisation of affixal inflectional morphology. To this end, three groups of participants were trained on plural inflectional suffixation of nouns in an artificial language, with each group being trained on a somewhat different list of words. Suffix type frequency was examined by including high, medium and low frequency suffixes (i.e. with a high, medium, or low number of trained words taking that suffix) in each group. To examine phonological predictability, lists of trained items were designed so that words' rimes provided cues to their inflection, with varying predictability across words and suffixes. Predictability was also manipulated between groups, with higher predictability in group C (where phonological cues fully predicted the affixes) than groups A and B. Suffix diversity (i.e. the number of rimes contained in the words that received each suffix) varied across suffixes within and between groups, allowing for its examination as a continuous variable.

We predicted that suffix type frequency and phonological predictability would affect the learning of trained words. In addition, we predicted that all three factors would affect the generalisation of the suffixes to untrained words. Effects of suffix type frequency and suffix diversity were examined on untrained words with and without rime cues. The following paragraphs summarise the effects found in the present study, and compare them to previous relevant research.

Suffix type frequency

Consistent with our predictions, suffix type frequency affected the learning of inflections for trained words. This was evident in better performance (more accurate and faster judgment) of trained items with high-frequency suffixes than those with medium- or low frequency suffixes, in groups A and B. Interestingly, however, trained items with low-frequency suffixes were learned as well or better than those with the medium-frequency suffix, suggesting that learning of these inflections may have been affected by other factors.

Suffix type frequency also affected the generalisation of inflections to new words. This was evident in the distribution of inflections for untrained words, which showed greater application of the high frequency suffix in the first session in untrained words both with and without rime cues. The critical role of suffix type frequency in inflecting untrained words with no phonological cues was also evident in the analysis of cosine similarities, showing greater reliance on suffix type frequency than on suffix diversity in the first session. Nevertheless, the reliance on suffix diversity increased by the 6th session, so that both factors played a similar role by the end of training. This suggests that the participants' sensitivity to suffix type frequency and their reliance on it in the generalisation to new words emerged rapidly, whereas reliance on other factors may emerge more gradually (see below). Note that based on participants' answers to the questionnaire in the 6th session, they had good explicit knowledge of suffixes' frequencies, but were not aware of applying this knowledge when inflecting untrained words, suggesting that the reliance on suffix frequency may be implicit.

Previous studies in natural languages have suggested that higher affix type frequency leads to better learning and generalisation of inflection in one's native language (Bybee, 1995; Dabrowska, 2004; Dabrowska & Szczerbiński, 2006; Köpcke & Wecker, 2017). The results of the current study show that this factor plays a role in learning morphological inflections – at least for plural inflections – in a second (artificial) language in adults as well. By using a controlled experimental design, the current study has also shown that reliance on inflections' frequencies is the strongest factor affecting generalization to new words (as compared to affix predictability and diversity), particularly in early stages of learning.

Suffix predictability based on phonological cues

Our results appear to indicate that participants became sensitive to the predictability of phonological cues in the stems, and applied this knowledge when inflecting trained words. Although accuracy in inflection of trained words may reflect item-specific memorization, the pattern of results suggests that participants also applied knowledge of phonological predictability of cues when inflecting trained words. Accuracy on trained words was highest for group C, in which suffixes had the highest phonological predictability (cues were deterministic), while groups A and B, which were trained on inflections with lower suffix predictability, showed lower accuracy. Additionally, the fact that inflections of words containing consistent rime cues were learned better than those of words containing inconsistent rime cues further suggests that the predictability of rime cues played a role in learning inflections of trained words. These results are consistent with previous findings showing that children rely on phonological cues for inflecting existing affixed forms in their first language (Laaha, 2011).

Our results also show that participants' sensitivity to the predictability of rime cues was generalized to untrained words containing such cues. First of all, as predicted, and consistent with the pattern for trained words, analysis revealed that group C produced the highest proportion of "optimal" responses to these words, and did so earlier than groups A and B. These results are consistent with previous studies showing that adults (Albright & Hayes, 2003) and children (Köpcke & Wecker, 2017) rely on phonological cues in generalizing inflectional affixes to novel words in their native language, as do children in the application of inflectional affixes to irregulars, in the form of over-regularizations (Hartshorne & Ullman, 2006; Marchman, 1997). Thus, our results demonstrate that the predictability of the affix based on phonological cues also plays a role in learning inflections in an (artificial) second language.

In addition, our results show that the reliance on phonological cues, as reflected in the application of optimal responses to untrained words, increased over the course of training, from the first to the sixth session, especially in groups A and B. This suggests that participants' reliance on suffix predictability to inflect novel words increases with experience, particularly when the cues are not deterministic. Finally, participants' answers on the questionnaire administered at the end of training indicate partial awareness of phonological cues embedded in trained items. The results also show more optimal responses among participants who reported noticing the phonological regularities, suggesting that the reliance on phonological cues was at least to some extent an explicit strategy.

Suffix diversity

Results of comparisons of cosine similarities between expected and actual distributions of inflections of untrained words without rime cues indicate an increase in reliance on suffix diversity from the first to the sixth session in groups A and B, but not in group C, where this reliance was already high at the end of the first session. This time course is very similar to the evolution of knowledge of phonological cues as evident in the reliance on suffix predictability. Both factors show slower emergence in groups A and B compared to group C, in which suffix predictability is higher. This supports the notion that the mental computation of diversity requires sensitivity to the phonological cues (i.e. predictability) on which this calculation depends, in this case rimes, a sensitivity which itself appears to have emerged gradually, as discussed above.

Reliance on suffix diversity may help explain the finding that participants applied the low-frequency suffixes as often as the high-frequency suffix to untrained words without rimes in the 6th session, since the low-frequency suffixes had the highest diversity when accumulated across all three suffixes. Note that suffix diversity, like suffix type frequency, is applicable to the inflection of any word, including those that do not share any cues with trained words. Diversity can therefore guide the selection of suffixes by default, when no information in the word itself guides the selection of the inflection (e.g. in untrained words with no phonological similarity to familiar words).

Overall, our results are consistent with previous studies showing a role for affixal diversity in the application of inflections in children's mother tongue (Dabrowska, 2004; Dabrowska & Szczerbiński, 2006), as well as with results of non-linguistic category learning studies (Hahn et al., 2005). Our results further demonstrate the role of affixal diversity in the generalization of plural inflection in adults learning a second (artificial) language, and show that the emergence of this effect depends on the evolving sensitivity to the predictability of phonological cues.

Implications

The results of the current study indicate that the learning of plural affixal inflections of trained words was affected by both suffix type frequency and suffix predictability, and that their generalization was modulated by all three factors examined: suffix type frequency, suffix predictability, and suffix diversity. This is broadly consistent with results of previous studies examining affixal inflection in natural first languages (e.g. Dabrowska, 2001, 2004; Dabrowska & Szczerbiński, 2006; Laaha, 2011; Laaha, Ravid, Korecky-Kröll, Laaha, & Dressler, 2006; Marchman, 1997). However, our results suggest that these factors may also affect learning of plural inflections in a second language, and demonstrate a possible time course for the emergence of these effects.

These three factors have also been found to affect learning and generalization in non-linguistic category learning studies (e.g. Barsalou, Huttenlocher, & Lamberts, 1998; Ell & Ashby, 2006; Hahn et al., 2005). If affixal plural inflection is affected by factors that also modulate learning and generalization in non-linguistic domains, as our results indicate, this suggests the involvement of domain-general learning mechanisms in morphology learning, such as those proposed by Ullman (2004, 2015; Hamrick, Lum, & Ullman, 2018), which ties morphology and other aspects of language to declarative and procedural memory. Indeed, category learning has been linked to both of these memory systems (Ashby & O'Brien, 2005).

Our results reveal what may be the emergence of default inflections, i.e. a preferential application of certain inflections for unfamiliar words which are phonologically dissimilar to familiar words. Our findings suggest that this depends both on the suffix type frequency and the suffix diversity of the inflection. Heavier reliance on suffix type frequency may resemble a "probability matching" strategy, in which affixes are used at the approximate frequencies observed in the input (Hudson Kam & Newport, 2009). In contrast, suffix diversity can lead to a deviation from the probabilities predicted by frequency alone. The existence of default inflections, and particularly putative "minority defaults," that is, low-frequency affixes that are applied by default to unfamiliar words (and whose default status cannot therefore be related to their frequency), has been taken to indicate reliance on knowledge of rules for performing inflections (Marcus et al., 1995). Marcus et al. (1995) note that the minority default –s in the German plural inflection system is characterised by the diverse circumstances in which it is applicable. It may therefore be worthwhile to examine more generally the extent to which the emergence of minority default inflections in natural languages is related to the inflections' diversity. This could be performed by comparing diversities of minority default inflections in various inflection systems, calculated using the formal definition proposed here, with diversities of other inflections within the same systems.

Although both suffix type frequency and suffix diversity appear to affect the emergence of default-like affixation, their relative influence seems to depend importantly on the amount of exposure to the language, as well as other properties of the language such as the consistency of phonological (and perhaps other) cues in predicting affixes. Learning suffix diversity requires the ability to classify the words receiving each inflection according to the cues they contain (i.e. it requires the formation of categories representing each of the cues) as a pre-requisite, whereas learning type frequency does not. This is reflected in the different time courses for the emergence of reliance on suffix type frequency and suffix diversity. Type frequency has a strong effect in early stages of learning, whereas reliance on diversity emerges more slowly, together with the evolving sensitivity to phonological cues. The fact that participants were able to acquire knowledge of phonological cues and suffix diversity following relatively limited exposure in the current study is not trivial. It implies that quite limited exposure may be sufficient for the acquisition of these two distinct, yet inter-dependent forms of knowledge: that of cues, which are properties of stems, and can be inferred by comparing stems and inflected forms, and that of diversities, which are properties of inflections, and require keeping track of cues present in words taking each inflection. Such a capacity to estimate distributions of phenomena, use these estimates to form categories, and then in turn estimate distributions across these newly formed categories, while simultaneously tracking distributions across multiple categories, can potentially enable elaborate classifications, and may therefore play a role in other learning processes as well.

The extent of predictability in the language seems to have an effect on the learning of relevant cues, and subsequently on the calculation of diversity, which depends on sensitivity to these cues. Thus, participants in group C, who trained on words in which rimes provided perfect (i.e. deterministic) cues, applied knowledge of rime cues, and also relied on diversity, to a greater extent than participants in both other groups, particularly at the end of the 1st session.

Second language learning

The current study was designed to model the learning of plural inflections in a second language. Accordingly, participants in the current study were adults (with ample prior experience in using plural inflections in their native language), and the words referred to familiar objects which correspond to words in their first language. Moreover, the design used in the current study, which involved the presentation of isolated nouns, followed by their inflection, is more akin to the explicit instruction of a second language than to the learning of inflections within the context of sentences with syntactic cues, typical of first language learning, or second language learning in immersion contexts. While the use of this design, and of an artificial language more generally, may raise concerns about the ecological validity of the study, it should be noted that similar designs have been employed in previous studies using artificial languages (e.g. Bybee & Newman, 1995; Ellis & Schmidt, 1997; 1998), as well as for examining generalization of inflections in natural languages, including by children (e.g. Berko, 1958; Köpcke & Wecker, 2017).

The results of the current study may therefore be more relevant to second language learning by adults than to children learning morphological inflections in their first language. Indeed, participants' responses to the questions presented to them at the end of the 6th session indicate they had good explicit knowledge of inflections' frequencies, and some explicit knowledge of the phonological cues embedded in the trained-items. Furthermore, explicit knowledge of phonological cues appears to have affected their generalization of inflections. These findings are in accordance with Ullman's declarative/procedural model, which predicts that adults learning a second language rely particularly on declarative memory for grammar, including for affixal inflection (Babcock et al., 2012; Hamrick et al., 2018; Ullman, 2001, 2015). Moreover, the model's prediction that, with exposure, rule-governed knowledge can be proceduralized, is consistent with the gradual emergence of the apparently more implicit knowledge of affixal diversity, and its role in the application of default-like affixes. The findings suggest that the examination of affixal diversity as a factor in proceduralization may be warranted in future studies. Note that although procedural-based grammar has been closely linked to rule-governed composition (Ullman, 2004; 2016), suggesting the possibility that the observed default-like affixation might involve compositional processes (see Introduction), the methods employed in the current study do not directly test for the storage vs. composition distinction, and thus this study does not specifically elucidate this distinction in L2. Nevertheless, the results of an fMRI study (Nevat, Ullman, Eviatar, & Bitan, 2017) that used the same artificial language and learning paradigm may shed light on this issue. These results show that by the end of training, only untrained affixed items activated frontal regions associated with composition, while trained affixed items, and especially those inflected with high-frequency suffixes, may have already been stored.

Although there are clear differences between children and adults, and between first and second language acquisition, the findings from the present study may nonetheless also be relevant to the child acquisition of affixal morphology, at least for plural inflection. Compared to adults, children's working memory and cognitive control processes are less developed (Thompson-Schill, Ramscar, & Chrysikou, 2009), which might help explain why children may possess less explicit knowledge of frequencies and phonological cues than adults, since explicit knowledge is linked to working memory (Ramscar & Gitcho, 2007; Ullman, 2016). Moreover, previous studies show that in their native language children have a stronger tendency to regularise using the most common forms, in conditions where adults reproduce a more veridical distribution of frequencies, suggesting a probability matching strategy (Hudson Kam & Newport, 2005; 2009). Nevertheless, children's performance on morphological inflections in their first and second language (Dabrowska, 2004; Dabrowska & Szczerbiński, 2006; Köpcke & Wecker, 2017; Laaha, 2011) suggests that they too rely on affix type frequency, phonological predictability, and phonological diversity, presumably in a more implicit form. Thus, the results from the present study may be at least indirectly relevant to children's acquisition of affixal morphology. Thus, it may be worthwhile to examine children's acquisition of the same or a similar artificial language as was examined here, which may reveal both differences and similarities between children and adults in the role of key distributional factors in the acquisition of affixal morphology.

Although our study examined plural affixal inflections, we suggest that our findings may be at least partly generalisable to other types of affixal inflections. As mentioned in the Introduction, nominal plural inflections appear to be easier and faster to acquire as compared to other inflections, perhaps because they often represent concrete lexical/semantic entities (Pinker, 1984).

However, other differences between plural inflection and other types of inflections may not be due to inherent properties of nominal plural inflection. For example, most differences found between English plural inflection (-s) and English verbal inflections (such as the third person singular -s or the past tense -ed) appear to be explained by properties of these specific English inflections – e.g. differences in type and token frequencies (Hsieh et al., 1999; Marchman, Plunkett, & Goodman, 1997), or phonological differences resulting from the position of the inflected form in the sentence (Hsieh et al., 1999) – rather than to inherent aspects of plural inflection itself. Thus, it does not seem unlikely that our findings from plural affixation may generalize, at least to some degree, to other types of affixal inflection. Nevertheless, the extent to which such generalization might hold remains unclear, and warrants clarification by future studies.

Finally, participants in our study were native speakers of Hebrew, a language that uses suffixes for plural inflection. In Hebrew, as in our artificial language, phonological cues to inflections are present at word ending positions (Berent et al., 1999; Ravid et al., 2008), and suffixes vary in how well they can be predicted from phonological cues. Various theories on second language learning suggest that properties of the L1 can crucially affect L2 acquisition (Bybee, 2008; Gass & Selinker, 2008; MacWhinney, 2001), suggesting that these similarities between the artificial language and participants' native language may have facilitated the learning of these properties. However, the effects of L1 on the acquisition of morphology more specifically have been debated (Clahsen et al., 2010; Koda, 2000; Luk & Shirai, 2009; Parodi, Schwartz, & Clahsen, 2004; Portin et al., 2008), so this remains to be tested in native speakers of languages with different morphological structures.

Conclusion

The current study demonstrates the potential importance of distributions of three factors to the learning and generalization of affixal plural inflections in adults learning a new language: suffix type frequency, phonological cue predictability and, importantly, suffix diversity. All three factors have also been found to affect learning in non-linguistic domains, suggesting that the underlying processes may be domain-general. While suffix type frequency is rapidly extracted in early stages of training and the knowledge is immediately applied to both trained and untrained words, the sensitivity to phonological cues emerges more slowly, and its application depends, at least to some extent, on explicit knowledge. Concurrently with the knowledge of phonological cues, participants gain sensitivity to suffixes' phonological diversity, and apply this sensitivity in inflecting new words with no phonological cues. As suffix type frequency and suffix diversity are properties of the suffix, they can guide the inflection of any new word, regardless of its properties. Reliance on these factors can therefore provide a possible explanation for the emergence of default inflections, at least for plural inflection. In particular, reliance on diversity could contribute to the emergence of minority default inflections. Future studies could examine the role of affixal diversity as formalised here in generalization in natural languages, in order to test whether and how affixal diversity, perhaps in combination with affix type frequency, might explain the emergence of default affixal inflection.

Acknowledgements

The authors wish to thank Dr. Irit Meir for her helpful advice during the design of the experiment and selection of stimuli. This study was funded by BSF grant number 2007077 to Tali Bitan and Michael T. Ullman.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by United States-Israel Binational Science Foundation [grant number 2007077].

References

Aduan-Mansour, J., & Bitan, T. (2017). The effect of stimulus variability on learning and generalization of reading in a novel script. *Journal of Speech Language and Hearing Research*. doi:10.1044/2017_JSLHR-L-16-0293

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90* (2), 119–161. doi:10.1016/S0010-0277(03)00146-X

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*(2), 83–89. doi:10.1016/j.tics.2004.12.003

Babcock, L., Stowe, J. C., Maloof, C. J., Brovetto, C., & Ullman, M. T. (2012). The storage and composition of inflected forms in adult-learned second language: A study of the influence of length of residence, age of arrival, sex, and other factors. *Bilingualism: Language and Cognition*, *15*(4), 820–840. doi:10.1017/S1366728912000053

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, *36*, 203–272. doi:10.1006/cogp.1998.0687

Berent, I., Pinker, S., & Shimron, J. (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition*, *72* (1), 1–44. doi:10.1016/S0010-0277(99)00027-X

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177. doi:10.1080/00437956.1958.11659661

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455. doi:10.1080/01690969508407111

Bybee, J. (2008). Usage based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–235). New York, NY: Routledge.

Bybee, J., & Newman, J. (1995). Are stem changes as natural as affixes? *Linguistics*, *33*, 633–654. doi:10.1515/ling.1995.33.4.633. doi:10.1093/oxfordhb/9780199738632.001.0001.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, *60*(1), 21–43. doi:10.1111/j.1467-9922.2009.00550.x

Dabrowska, E. (2001). Learning a morphological system without a default: The Polish genitive. *Journal of Child Language*, *28*(3), 545–574. doi:10.1017/S0305000901004767

Dabrowska, E. (2004). Rules or schemas? Evidence from Polish. *Language and Cognitive Processes*, *19*(2), 225–271. doi:10.1080/01690960344000170

Dabrowska, E., & Szczerbiński, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language*, *33*(3), 559–597. doi:10.1017/S0305000906007471

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*, 344–358. doi:10.1016/j.jml.2011.01.003

Ell, S., & Ashby, F. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception & Psychophysics*, *68*(6), 1013–1026. doi:10.3758/BF03193362

Ellis, N., & Schmidt, R. (1997). Language and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*, 27.

Ellis, N., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*(2/3), 307–336. doi:10.1080/016909698386546

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. M. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*, 822–847. doi:10.1111/cogs.12257

Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Filipović Đurđević, D., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, *13*(2), 119–135. doi:10.1017/S1366728909990459

Field, A. (2005). *Statistics using SPSS* (2nd ed.). London: Sage.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). New York, NY: Routledge. Retrieved from https://doi.org/10.1057/9780230240780_6

Gerken, L. A. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*, B67–B74. doi:10.1016/j.cognition.2005.03.003

Hahn, U., Bailey, T., & Elvin, L. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302. doi:10.3758/BF03195318

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1713975115

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say "holded" more than boys. *Developmental Science*, *9*(1), 21–32. doi:10.1111/j.1467-7687.2005.00459.x

Hsieh, L., Leonard, L. B., & Swanson, L. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language*, *26*(3), 531–543. doi:10.1017/S030500099900392X

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195. doi:10.1207/s15473341lld0102_3

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66. doi:10.1016/j.cogpsych.2009.01.001

Jia, G. (2003). The acquisition of the English plural morpheme by native mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1297–1311. doi:10.1044/1092-4388(2003/101

Kaivapalu, A., & Martin, M. (2007). Morphology in transition: Plural inflection of Finnish nouns by Estonian and Russian learners. *Acta Linguistica Hungarica*, *54*(2), 129–156. doi:10.1556/ALing.54.2007.2.2

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*(4), 703–746. doi:10.1111/j.1467-9922.2008.00477.x

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. *Applied Psycholinguistics*, *21*(3), 297–320.

Köpcke, K. (1998). The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language*, *25*(2), 293–319. doi:10.1017/S0305000998003407

Köpcke, K. M., & Wecker, V. (2017). Source- and product-oriented strategies in L2 acquisition of German plural marking. *Morphology*, *27*, 77–103. doi:10.1007/s11525-016-9292-5

Laaha, S. (2011). Sonority, gender and the impact of suffix predictability on the acquisition of German noun plurals. *Language, Interaction and Acquisition*, *2*(1), 82–100. doi:10.1075/lia.2.1.04laa

Laaha, S., Ravid, D., Korecky-Kröll, K., Laaha, G., & Dressler, W. (2006). Early noun plurals in German: Regularity, productivity or default? *Journal of Child Language*, *33*(2), 271–302. doi:10.1017/S0305000906007379

Luk, Z. P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural-*s*, articles, and possessive's. *Language Learning*, *59*(4), 721–754. doi:10.1111/j.1467-9922.2009.00524.x

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69–90). Cambridge:

University Press. Retrieved from https://doi.org/10.1017/CBO9781139524780

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology and neighborhood structure. *Cognitive Science*, *21*(3), 283–304. doi:10.1207/s15516709cog2103_2

Marchman, V. A., Plunkett, K., & Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language*, *24*(3), 767–779. Retrieved from https://ora.ox.ac.uk/objects/uuid:a377fdd2-b3dc-4c89-897c-7bb9124a97ac

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*(3), 189–256. doi:10.1006/cogp.1995.1015

Maslen, R., Theakston, A., Lieven, E., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1319–1333. doi:10.1044/1092-4388(2004/099

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–947. doi:10.1162/jocn_a_00119

Morgan-Short, K., & Ullman, M. T. (2012). The neurocognition of second language. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition*. New York, NY: Routledge.

Nevat, M., Ullman, M. T., Eviatar, Z., & Bitan, T. (2017). The neural bases of learning and generalization of morphological inflection. *Neuropsychologia*, *98*, 138–155. doi:10.1016/j.neuropsychologia.2016.08.026

Parodi, T., Schwartz, B. D., & Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics*, *42*(3), 669–705. doi:10.1515/ling.2004.022

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1991). Rules of language. *Science*, *259*(5019), 530–535. doi:10.1126/science.1857983

Portin, M., Lehtonen, M., Harrer, G., Wande, E., Niemi, J., & Laine, M. (2008). L1 effects on the processing of inflected nouns in L2. *Acta Psychologica*, *128*, 452–465. doi:10.1016/j.actpsy.2007.07.003

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*(1), 1–56. doi:10.1080/01690969308406948

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, *11*(7), 274–279. doi:10.1016/j.tics.2007.05.007

Ravid, D., Dressler, W. U., Nir-Sagiv, B., Korecky-Kröll, K., Souman, A., Rehfeldt, K., … Gillis, S. (2008). Core morphology in child directed speech: Crosslinguistic corpus analyses of noun plurals. In H. Behrens (Ed.), *Corpora in language acquisition research: Finding structure in data* (pp. 25–60). Amsterdam: Benjamins. Retrieved from https://doi.org/10.1075/tilar.6.05rav

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*, 30–54. doi:10.1016/j.cogpsych.2012.09.001

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland & D. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.

Salton, G., & Lesk, M. E. (1968). Computer evaluation of indexing and text processing. *Journal of the Association for Computer Machinery*, *15*(1), 8–36.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260. doi:10.1017/S1366728908003404

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112. doi:10.1016/j.cognition.2012.06.014

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *18*(5), 259–263. doi:10.1111/j.1467-8721.2009.01648.x

Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space and time. A review of the role of cross-language similarity in L2 (morpho)syntactic processing as revealed by fMRI and ERP methods. *Studies in Second Language Acquisition*, *33*(1), 91–125. doi:10.1017/S0272263110000549

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*(1), 105–122. doi:10.1017/s1366728901000220

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–270. doi:10.1016/j.cognition.2003.10.008

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok & S. A. Small (Eds.), *The neurobiology of language* (pp. 953–968). San Diego, CA,USA: Elsevier.

Ullman, M.T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 135–158). New York, NY: Routledge.

Vender, M., Mantione, F., Savazzi, S., Delfitto, D., & Melloni, C. (2017). Inflectional morphology and dyslexia: Italian children's performance in a nonword pluralization task. *Annals of Dyslexia*, *67*(3), 401–426. doi:10.1007/s11881-017-0152-8

Wood, J. N., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology*, *45*(1), 202–206. doi:10.1037/a0014432