Journal of Experimental Psychology: Learning, Memory, and Cognition

The Influence of Type and Token Frequency on the Acquisition of Affixation Patterns: Implications for Language Processing

Ansgar D. Endress, and Marc D. Hauser

Online First Publication, August 30, 2010. doi: 10.1037/a0020210

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/10/$12.00 DOI: 10.1037/a0020210

The Influence of Type and Token Frequency on the Acquisition of Affixation Patterns: Implications for Language Processing

Ansgar D. Endress and Marc D. Hauser
Harvard University

Rules, and exceptions to such rules, are ubiquitous in many domains, including language. Here we used simple artificial grammars to investigate the influence of 2 factors on the acquisition of rules and their exceptions, namely type frequency (the relative numbers of different exceptions to different regular items) and token frequency (the number of exception tokens relative to the number of regular tokens). We familiarized participants to either a prefixation pattern (where regulars started with /ZaI/ and exceptions ended with /ZaI/) or a suffixation pattern (where regulars ended with /ZaI/ and exceptions started with /ZaI/). We show that the type and the token frequency of regular items and exceptions influence in different ways what participants can learn. For the exceptions to be learned, they have to occur sufficiently often so that participants can memorize them; this can be achieved by a high token frequency. However, a high token frequency of the exceptions also impaired the acquisition of the *regular* pattern. In contrast, the type frequency of the patterns seemed to determine whether the regular pattern could be learned: When the type frequency of the regular items was sufficiently high, participants successfully learned the regular pattern even when the exceptions were played so often that 66% of the familiarization items were exceptions. We discuss these findings in the context of general learning mechanisms and the role they may play in language acquisition.

Keywords: language acquisition, morphology, type frequency, token frequency, distributional learning

Supplemental materials: http://dx.doi.org/10.1037/a0020210.supp

Rules, and exceptions to such rules, occur in many cognitive domains, from the general problem of categorization (e.g., Erickson & Kruschke, 1998; Needham, Dueker, & Lockhead, 2005; Nosofsky, Palmeri, & McKinley, 1994; Palmeri & Nosofsky, 1995; see also Quinn & Bhatt, 2005) to the more specific problem of reading (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Miceli, Capasso, & Caramazza, 1994). The domain of language is perhaps the most obvious context, and a particularly prominent example is inflectional morphology. Many inflectional systems comprise both default rules (such as the /-ed/ suffix for the English past tense) and exceptions to such defaults (forms such as *go–went*). Though this observation is uncontroversial, different researchers have proposed different theories regarding how rules and exceptions are processed. These theories range from proposals that all inflected forms are processed by rule-based mechanisms (e.g., Kiparsky, 1982; Taft, 1979; Taft & Forster, 1975) to proposals that regulars are processed by rule-based mechanisms and exceptions by more associative mechanisms (e.g., Caramazza, Laudanna, & Romani, 1988; Caramazza, Miceli, Silveri, & Laudanna, 1985; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Marslen-Wilson & Tyler, 1997, 1998; Pinker, 1991, 1999; Pinker & Prince, 1988; Tyler, Marslen-Wilson, & Stamatakis, 2005), to proposals that all forms are processed by general associative machinery (e.g., Eddington, 2009; Hahn & Nakisa, 2000; Hare & Elman, 1995; Hare, Elman, & Daugherty, 1995; Joanisse & Seidenberg, 1999; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986; Seidenberg, 1997; Westermann & Plunkett, 2007) or memory mechanisms (e.g., Butterworth, 1983). However, except for corpus studies (e.g., Bybee, 1995; Clahsen, Rothweiler, Woest, & Marcus, 1992; Dabrowska, 2001; Dabrowska & Szczerbiński, 2006; Köpcke, 1998; Marcus et al., 1995) and neural network simulations (e.g., Hahn & Nakisa, 2000; Hare et al., 1995; Plunkett & Marchman, 1991), the factors determining the acquisition of rules and exceptions have rarely been investigated experimentally, at least in the domain of language.

Here we examine this issue by exploring how human adults acquire a simple system of rules and exceptions involving prefixation and suffixation patterns. Both prefixation and suffixation play important roles in many inflectional systems. There are more than 4 times as many languages predominantly using suffixes for tense marking compared with the languages predominantly using prefixes, and, more generally, suffixation is much more common across the world's languages than prefixation (Dryer, 2005). Still, many languages do use prefixation. For example, in classical Greek, different past tenses (e.g., imperfect and aorist) use an

Ansgar D. Endress and Marc D. Hauser, Department of Psychology, Harvard University.

Funding for this work was provided by Harvard Mind, Brain, and Behavior Initiative grants to Marc D. Hauser, Andrew Nevins, and Ansgar D. Endress, as well as gifts from Jeffrey Epstein and Stanley Shuman. We thank Pierre Perruchet and Harald Baayen for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Ansgar D. Endress, who is now at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Room 46-4127, Cambridge, MA 02139. E-mail: ansgar.endress@m4x.org

/ɛ̀-/prefix to mark the past (as well as suffixes for person and number agreement); Malagasy, the language spoken in Madagascar, uses the /n-/ prefix as a past tense marker (Garvey, 1964); and the German past participle combines the /ge-/ prefix with one of two suffixes.

Although prefixation and suffixation patterns map onto important aspects of linguistic structure, and were targeted for this reason, we leave to the General Discussion the question of whether the underlying mechanisms are specific to linguistic processing or more domain-general. In the specific experiments, participants listened to a number of words most of which conformed to a common pattern. For example, most words shared the same prefix, but some exceptions occurred with the prefix syllable placed in the suffix position. We asked how well participants would apply the overall, default pattern to words they had heard and to new items they had not heard, and how well they would learn the exceptions. We measured the participants' performance as a function of two orthogonally manipulated factors: the number of exceptions and the (token) frequency of these exceptions.

Effects of Type and Token Frequency

Frequency effects are one of the oldest and most robust phenomena in psycholinguistic experimentation, typically yielding processing advantages for more frequent words relative to less frequent words (e.g., Cattell, 1886; Forster & Chambers, 1973; Solomon & Postman, 1952). However, frequencies can be measured in different ways. Take syllable frequencies as an example. On the one hand, one can search all words in a corpus that share a given initial syllable and sum up the total number of occurrences of these words. This would yield a measure of the *token* frequency of that syllable. On the other hand, one can count the number of different words in the corpus that share a given initial syllable. This would give a measure of the *type* frequency of that syllable. Conrad, Carreiras, and Jacobs (2008) showed that reaction times in a lexical decision task are slower for words whose initial syllable has high token frequency than for words whose initial syllable has low token frequency. In contrast, reaction times are faster for words whose initial syllable has high type frequency than for words with a low-type-frequency first syllable. These results suggest that frequency can be measured in different ways and that type and token frequency can sometimes have opposite effects on speech processing.

Note that type and token frequency do not necessarily have opposite effects in language processing. Most relevant to the current article, type and token frequency typically both have facilitatory effects on morphological processing (Balling & Baayen 2008; Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009). Here, in contrast, we studied the effects of type and token frequency on how inflection patterns are learned (rather than processed once they are acquired) and, as the Conrad et al. (2008) studies showed in the case of syllable frequencies, how these effects might well be different.

Previous artificial language studies have revealed that the "frequency" of linguistic patterns influences how learners acquire them. For example, Hudson Kam and Newport (2005) claimed that, when grammarlike regularities are made "inconsistent" by introducing "exceptions" to the regularities, adults learn both the overall regularities and the exceptions, and children tend to make the exceptions regular (see also Hudson Kam & Newport, 2009, for similar claims). With much younger infants, inconsistent input seems to make learning such regularities harder (Gómez & Lakusta, 2004). Other authors have suggested that regularities need to be instantiated by some minimal number of examples to be learnable; if there are fewer distinct examples of a regularity, infants have difficulty learning it (Gerken & Bollt, 2008; see also Gómez, 2002). However, in these experiments, type and token frequency have not been varied independently. As a result, the specific contributions of these forms of frequency are presently unknown.

In the case of natural language morphology, the role of type and token frequency on productivity in inflectional morphology has long been controversial. Productivity is the extent to which a morphological pattern can apply to novel words. English children, for instance, judge that the plural of a nonsense noun such as *wug* is *wugs* (Berko, 1958); hence, the English plural *s* is productive. Different inflectional patterns differ in their productivity (e.g., Bybee & Moder, 1983; Prasada & Pinker, 1993). For example, though English speakers are willing to extend the past tense formation of verbs such as *swing–swung* to novel nonsense words, they do so only if the novel word is sufficiently similar to *swing* (e.g., *spling–splung*). In contrast, they readily extend the regular past tense formation (e.g., *walk–walked*) to new words even if they do not resemble known verbs. The regular past tense thus is much more productive than formations such as *swing–swung*, *sink–sank*, and so on.

Although it is uncontroversial that different morphological formations have different productivities, the origin of these differences is still debated, and has been used to draw strong conclusions about the fundamental nature of the mechanisms involved in morphology. Authors who have argued that type frequency determines productivity have typically concluded that morphology relies on statistical computations, whereas authors arguing against this view have typically favored symbolic, rule-based morphological mechanisms. In the General Discussion, we return to the question of whether these assumptions are justified. For example, Marcus et al. (1995) suggested that the German default past participle has a lower type frequency than the nondefault formation (i.e., a lower number of different words takes the default), concluding that the default plural formation must rely on a symbolic, rule-based mechanism. Bybee (1995), in contrast, showed that this conclusion strongly depends on how the token frequency is counted. Whereas Marcus et al. (1995) considered only the 1,000 most frequent words to calculate the type frequencies, Bybee (1995) used a more complete corpus and found the opposite: That is, the default pattern is favored by the type frequency.

There have been other attempts to clarify the contributions of type and token frequencies on the acquisition of morphological patterns. For example, Nicoladis, Palmer, and Marentette (2007) compared the performance of monolingual and bilingual children in inflectional morphology. They observed that bilinguals are exposed half as much to each of their languages as monolinguals (because they hear two languages, and the total language input is presumably constant for monolinguals and bilinguals). As a result, so they reasoned, this difference should affect just the token frequencies of words in the inputs to monolinguals and bilinguals, but not the type frequencies of different inflectional patterns. This assumption, however, is not necessarily licensed, as bilingual children know a comparable number of words to monolinguals when both languages are taken together but fewer words in each of their languages (e.g., Pearson, Fernández, & Oller, 1993; Umbel, Pearson, Fernández, & Oller, 1992). This may, in turn, affect the type frequencies of different inflectional patterns. In the absence of detailed analyses of the input, it is thus unclear whether bilingual language acquisition can unravel the respective contributions of type and token frequencies for the acquisition of inflectional patterns.

Dabrowska and Szczerbiński (2006) investigated the contributions of type and token frequency to the productivity of different inflectional patterns by correlating Polish-born children's use of different inflectional patterns with the type frequency of these patterns in Polish. They found a strong correlation between the type frequency of the patterns and their productivity, concluding that infants acquire morphological patterns through "usage-based" mechanisms. However, the type frequency of the patterns was also strongly correlated with their token frequency, making strong conclusions about the relative contributions of type and token frequencies difficult. This reflects again the difficulty of controlling for the exact properties of the input when working with realistic corpora, and of disentangling the relative contributions of the many correlated variables that can be used to describe corpora. Here we attempted to isolate the role of token frequency using an artificial language learning paradigm, providing precise control over the properties of the language as a result of our specific design.[1]

In contrast to type frequency, there seems to be an agreement that high token frequency counts help people memorize words (e.g., Bybee, 1995; Marcus et al., 1995; Pinker, 1991; Plunkett & Marchman, 1993; Tomasello, 2000). The general assumption seems to be that words (and patterns) are memorized that occur often enough. The agreement thus seems to concern the *absolute* number of occurrences of a word and not its frequency *relative* to other words. However, as this assumption has, to our knowledge, never been tested empirically, we also tease apart the influence of the (a) absolute number of occurrences of items and (b) the relative number of occurrences of these items relative to others.

Of course, there are other factors that influence morphological processing and that might, therefore, influence the acquisition of morphological patterns as well, for example, the semantics of a word (Baayen & Moscoso del Prado Martín, 2005; Ramscar, 2002), its historical status (i.e., whether or not it is a loan word; Keuleers et al., 2007), and the saliency of the affixes (Järvikivi, Bertram, & Niemi, 2006). Our experiments allowed us to abstract away from such factors by employing a learning situation that does not involve any meaning or different affixes, targeting specifically the contributions of type and token frequency on the acquisition of inflectional patterns.

The Current Experiments

In the experiments presented below, we investigated the respective contributions of type and token frequency to the learning of two very simple "inflectional" patterns: prefixation and suffixation. Specifically, we exposed participants to monosyllabic nonsense words (hereafter called "stems"). For half the participants, most of the stems had the same prefix (i.e., an initial syllable), whereas some exceptions had the same syllable as a suffix but not as a prefix. For the remaining participants, the roles of prefixation and suffixation were reversed. We asked how the relative type and token frequencies would influence the learning and generalization of these patterns. Specifically, after exposure to the regular items and the exceptions, participants were presented with regular stems they had heard, with the exceptions, and with new stems (hereafter called "holdouts"), and had to judge whether these stems were more likely to take a prefix or a suffix.

We chose prefixation and suffixation patterns because they appear particularly simple and may be based on basic perceptual or memory mechanisms. Indeed, across the languages of the world, adding an affix in the first position (e.g., *re-do*) or the last position (e.g., *walk-ed*) is much more frequent than adding affixes in other positions (Greenberg, 1957; Julien, 2002). This observation correlates with results in serial memory tasks showing that the first and the last position in a sequence are precisely the positions that can be encoded accurately (see, e.g., Henson, 1998, 1999, for reviews). On the basis of this correlation, and the observation that edge positions are also encoded much better than nonedge positions in artificial grammar learning experiments (e.g., Endress & Mehler, 2009, 2010; Endress, Scholl, & Mehler, 2005), it has been suggested that affixation patterns in natural languages may rely on a similar basic positional memory mechanism (Endress, Nespor, & Mehler, 2009); in line with this view, cotton-top tamarin monkeys readily learn such patterns (Endress, Cahill, Block, Watumull, & Hauser, 2009). Irrespective of whether these suggestions turn out to be true, prefixation and suffixation seem to be patterns that are particularly easy to learn, across stimulus presentation, modality, and species.

In the following descriptions of the experiments, we focus on participants for whom the majority of stems obeyed a prefixation pattern; for ease of exposition, we do not mention the remaining participants, for whom the roles of prefixation and suffixation were simply inverted. The experiments are summarized in Table 1. In Experiment 1, we asked whether participants could learn a basic affixation pattern when no exceptions were present. In Experiment 2, we introduced four exception stems out of a total of 30 stems. For example, when most stems were prefixed with the same syllable, these exceptions were suffixed with that syllable. In Experiments 3 and 4, we investigated the role of token frequency for the learning of the exceptions by presenting the exception stems 5 and 14 times as often as the regular stems, respectively.

In Experiment 5, we investigated the role of the type frequency on the learning of the regular items and the exceptions. In that experiment, we matched the token frequency of the exceptions to that in Experiment 4; however, we halved the type frequency of the exceptions by including only two exceptions (out of 30 items). Experiment 6 controlled for alternative interpretations of Experiment 5.

Finally, the goal of Experiments 7 and 8 was to disentangle the two possible roles of the token frequency on the learning of the exceptions and the regular items. Recall that exceptions with high token frequencies may be learned particularly well because they

[1] Natural corpora also allow for alternative measures of the type frequency of a word (e.g., its morphological family size; see, e.g., Schreuder & Baayen, 1997). In our experiments, in contrast, such measures are not defined, as we do not have any morphological families.

Table 1
Summary of the Experiments

Experiment	Number of exceptions	Frequency (%)		Number of occurrences	Trials, % correct		
		Type	Token		Old	Holdout	Exception
1	0	0.0	0.0	0	92.7***	90.23***	N/A
2	4	13.3	13.3	3	79.17***	74.74***	36.46†
3	4	13.3	43.5	15	57.29	56.25	58.33
4	4	13.3	68.3	42	60.16	53.11	71.88**
4[a]	4	13.3	68.3	42	55.73	53.65	71.88**
5	2	6.7	66.7	84	70.83**	66.4*	76.04***
5[a]	2	6.7	66.7	84	74.48***	72.66**	70.83**
6	4	13.3	81.16	84	56.25	47.50	67.50*
7	2	6.7	33.3	21	73.12*	67.19*	67.5*
8	2	6.7	6.7	21	76.88	85.0***	63.75

Note. N/A = not applicable.
[a] Replication.
† $p = .06$. * $p < .05$. ** $p < .01$. *** $p < .001$.

occur often (in absolute terms) or occur more often relative to other items. In Experiment 8, the absolute number of occurrences of the exceptions was the same as in Experiment 7; however, the exceptions occurred as often as the regular items such that they were no longer frequent relative to the regulars. In Experiment 7, in contrast, the exceptions occurred as often as in Experiment 8, but the regular items occurred much less frequently. If a high relative token frequency has an influence on learning the exceptions over and above their absolute number of occurrences, we should observe a difference between these conditions.

Given the long-standing debates about the mechanisms involved in inflectional morphology, and the importance of type frequencies for views about the nature of these mechanisms, we note that our experiments do not rely on any particular assumptions about the nature of the mechanisms involved in morphology, the relevant learning mechanisms, or the domain of representation. That is, we present our experiments intentionally in a theory-neutral way, because we believe that our results can be modeled by both symbolic and associationist models. As a result, the only assumption we make is that people have the necessary machinery to learn and represent such patterns. In the General Discussion, we discuss the implications of our results for the underlying mechanisms.

Experiment 1: Learning a Basic Affixation Pattern With No Exceptions

In Experiment 1, we asked whether participants could learn basic affixation patterns in the absence of exceptions. Participants were presented with "affixed" stems. For participants in the prefixation group, the stems were preceded by the syllable /ZaI/; for participants in the suffixation group, this syllable followed the stems. Following this familiarization phase, we presented participants with old stems and new stems (holdouts), asking them to decide whether these were more likely to be prefixed or suffixed.

Method

Participants. Sixteen participants (10 women, six men; mean age = 23.8 years; range: 18–41) took part in this experiment for course credit or monetary compensation. In this and all other experiments, participants were recruited through the Harvard University study pool, were native speakers of American English, and reported normal or corrected-to-normal vision with no known auditory impairment. Each participant took part in only one experiment.

Apparatus. The experiment was run with PsyScope X (http://psy.ck.sissa.it). Stimuli were presented over headphones, and responses were collected on premarked keys on a keyboard.

Materials. In this and all other experiments, stimuli were created with the en1 voice (male British English) of Mbrola (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). Vowels had a duration of 200 ms. The consonant duration was 75 ms for those that occurred in consonant clusters and 150 ms for single consonant onsets or codas. The fundamental frequency was 150 Hz. There was no silence between stems and affixes. Sound files were mono with a sample frequency of 16 kHz and a sample width of 16 bits. Files were produced in the AU file format with Mbrola and then converted to AIFF with SoX (http://sox.sourceforge.net).

The affix was always /ZaI/. The stems (see below) were chosen to be phonetically variable and not to resemble real English words. (All familiarization stems are shown in Table S.1 in the supplementary online material.) Consequently, some of these stems had low (but clearly nonzero) phonotactic probabilities in English. To make sure that these items could be processed by English speakers, we presented a naive native speaker of American English with each of these items and asked her to repeat them. Though she agreed that some of the words were atypical for English, she had no difficulty in repeating all the items. Hence, we assumed that our stimuli should be readily processed by native speakers of American English.

Familiarization. Participants were told that they were to hear Martian words and were instructed to listen to them carefully. Participants in the prefixation group then heard all stems preceded by the affix /ZaI/; participants in the suffixation group heard all stems followed by /ZaI/. The 30 words were played three times, leading to a total familiarization of 90 words. Words were played in different random orders for different participants, with the constraint that words could not be repeated. Words were separated by a silence of 1 s.

Test. Following familiarization, participants were informed that the Martian words they just heard conformed to a regular pattern. Then they were presented with pairs of items and had to decide which of the items was more likely to be a correct Martian word. Both items in each pair shared the stem syllable, but one was prefixed and one suffixed. There was a silence of 1 s between the items in a trial; the next trial started 1 s after the participant's response.

In this experiment, there were two types of test trials. We used 12 old stems that participants had encountered during familiarization and 16 new stems (holdouts) that participants had not heard during familiarization, yielding 28 test trials in total. In both trial types, half the trials started with a prefixed item and half with a suffixed item. (All stems used during test are listed in Table S.2 in the supplementary online material.)

Scoring. In this and all other experiments, we scored a response in an old or holdout trial as correct if the participant selected the affixation pattern of the majority of the words during familiarization (i.e., prefixation for the prefixation group and suffixation for the suffixation group). In experiments with exception trials, we scored a response as correct if the participant selected the affixation pattern that the exception showed during familiarization.

In the analyses below, we primarily analyzed each trial type separately. We compared trial types only when we had predictions for these comparisons. Unless otherwise stated, one-sample tests are based on the Wilcoxon signed rank test and two-sample tests on the Mann–Whitney *U* test, which was additionally evaluated with a permutation test based on the *t* statistic using 10,000 random samples. Analyses of variance (ANOVAs) were performed on rank-transformed data, which we evaluated using both the *p* values as determined from the *F* ratios and the permutation tests based on these *F* ratios, again with 10,000 random samples. Although we do not have a sufficient number of items to perform item-by-item analyses with the other statistical tests, the mixed-effects model analyses described below include the specific items as random factors.

We performed several between-experiment comparisons using logistic mixed-effects models with fixed-effect predictors as specified below (see Baayen, Davidson, & Bates, 2008, for an informal introduction to such models). Mixed models are "mixed" because they simultaneously include fixed factors and random factors. In all mixed-effects analyses, we started with random intercepts for participants, stems, and trial number, but used a likelihood ratio test to keep only those random effects that significantly contributed to the likelihood of the models. We then evaluated the significance of the various estimates using *Z* tests.

All statistical tests reported in the manuscript are two-tailed with a chance level of 50% and a significance threshold of 0.05.

Results and Discussion

As shown in Figure 1, participants were almost perfect on the old trials (percent correct: $M = 92.7\%$, $SD = 9.0\%$; $W = 136$, $p < .0004$, Cohen's $d = 4.7$, 95% CI [87.88%, 97.54%]). There was no difference between the two language groups ($U = 29$, $p = .78$, *ns*). Likewise, participants were almost perfect on the holdout trials ($M = 90.23\%$, $SD = 14.8\%$; $W = 120$, $p = .0005$, Cohen's $d = 2.7$, 95% CI [82.36%, 98.11%]), with no difference between the two language groups (i.e., prefixation vs. suffixation; $U = 31.5$, $p > .99$, *ns*).

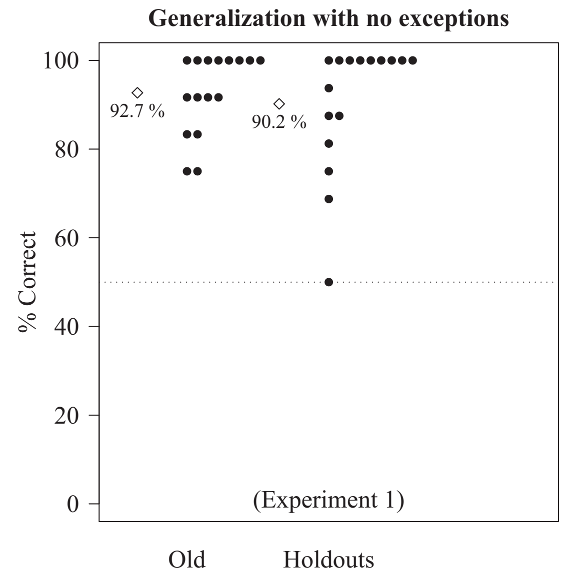

Figure 1. Results of Experiment 1. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When no exceptions were present, participants performed almost perfectly on both the old trials and the holdout trials.

Unsurprisingly, therefore, participants had no problem learning a basic affixation pattern. In Experiment 2, we asked whether participants can maintain this level of performance in the face of a number of exception stems that failed to conform to the overall affixation pattern.

Experiment 2: Learning an Affixation Pattern With Exceptions

In Experiment 2, we asked whether participants could learn an affixation pattern in the face of some exceptions. Specifically, four of the 30 familiarization stems now took the opposite affixation pattern from the majority of words (i.e., suffixation in the prefixation group and prefixation in the suffixation group). Subsequently, we tested participants on old regular stems, new stems not heard during familiarization, and the exceptions.

Method

Participants. Twenty-four new native speakers of American English (15 women, nine men; mean age = 21.4 years; range: 18–37) from the same population as in Experiment 1 took part in this experiment.

Familiarization. The familiarization was identical to that of Experiment 1 except that four stems (the exceptions) took the opposite affixation pattern from the majority of stems. These stems were /gA/, /prəʊ/, /smi/, and /kweɪb/. Words were presented in different random orders for different participants, with the constraint that words could not be repeated and that no more than three regular items or exceptions could occur in a row.

Test. The test phase was similar to that of Experiment 1 except that participants were presented with exception trials in addition to the old and holdout trials. In exception trials, participants were presented with the exception stems and had to decide whether they should be prefixed or suffixed.

The exception stems were among those used for the old trials in Experiment 1. In total, participants thus completed eight old trials, 16 holdout trials, and four exception trials. These trials were presented in random order. In each trial type, half the trials started with a prefixed item and half with a suffixed item.

Results

As shown in Figure 2, participants performed above chance in the old trials (M = 79.17%, SD = 20.7%; W = 207.5, p = .0001, Cohen's d = 1.4, 95% CI [70.41%, 87.93%]) and in the holdout trials (M = 74.74%, SD = 23.3%; W = 236, p = .0004, Cohen's d = 1.1, 95% CI [64.88%, 84.6%]). In contrast, they tended to perform below chance in the exception trials (M = 36.46%, SD = 32.1%; W = 43, p = .06, Cohen's d = 0.42, 95% CI [22.90%, 50.02%]).

An ANOVA on ranks with language (prefixation vs. suffixation default) as between-subjects factor and trial type (old vs. holdout vs. exception) as within-subjects factor, revealed a main effect of trial type, $F(2, 44) = 15.9$, $p < .0001$, $p_{\text{perm}} < .00001$, $\eta_p^2 = .416$, but no other main effects or interactions. Whereas the performance on old trials did not differ from that on holdout trials (U = 118, p = .156, *ns*), the performance on exception trials was lower than on old trials (U = 251.5, p < .0001, Cohen's d = 1.0, 95% $\text{CI}_{\text{median diff}}$ [25.0%, 62.5%]) and on holdout trials (U = 247.5, p < .0001, Cohen's d = 0.9, 95% $\text{CI}_{\text{median diff}}$ [21.9%, 56.3%]).

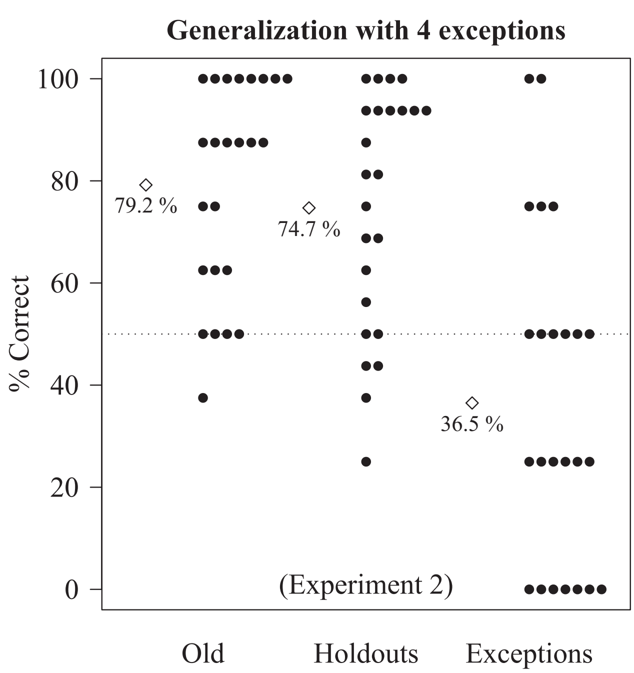

Figure 2. Results of Experiment 2. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When four out of 30 familiarization items were exceptions, participants learned the majority pattern, but also extended this pattern to the exceptions.

We compared Experiments 1 and 2 using a logistic mixed-effects model with the following fixed-effect predictors: language, trial type (excluding exceptions, which were not present in Experiment 1), experiment, Trial Type × Language, Trial Type × Experiment. Participants and trial number were used as random effects. (As mentioned above, the original model specification included a random intercept for stems, which turned out not to contribute significantly to the likelihood of the model.) We observed a significant effect of experiment (Z = 2.48, p = .013) but no other main effects or interactions (see Table S.3.1 in the supplementary online material for more detailed results). Accordingly, participants performed worse in Experiment 2 on both old trials (U = 262, p = .047) and holdout trials (U = 278.5, p = .015).

Discussion

Whereas participants in Experiment 1 were exposed to a perfectly consistent affixation pattern with no exceptions, they were introduced to some exceptions in Experiment 2. Results showed that participants failed to learn the exceptions. As each word was presented only three times, participants may not have heard the exceptions often enough to remember them. Accordingly, participants may have overgeneralized the majority pattern to the exceptions.

Remarkably, even though participants failed to learn the exceptions, their presence decreased performance on both old and holdout trials. Specifically, when only 13.3% of the stems were exceptions (by both type and token counts) and participants failed to learn the exceptions, their learning of the default pattern was also impaired.

Given that our level of exposure to the exceptions was insufficient for learning in Experiment 2, we designed two follow-up experiments in which we increased the token frequency. In Experiments 3 and 4, we presented the exceptions five and 14 times, respectively, more often than the regular items. As a result, these manipulations maintained a constant type frequency (because we did not change the number of exceptions) but increased the token frequency to 43.5% and 68.3%, respectively (measured as the number of exception tokens divided by the total number of tokens); the number of occurrences increased to 15 in Experiment 2 and 42 in Experiment 4.

Experiment 3: Learning an Affixation Pattern With Fivefold the Frequency of Exceptions

Method

Experiment 3 was identical to Experiment 2 except that the exceptions were presented 15 times (rather than three times as in Experiment 2). We tested 24 new native speakers of English (17 women, seven men; mean age = 21.5 years; range: 15–31).

Results and Discussion

As shown in Figure 3, the participants' performance did not differ from chance in old trials (M = 57.29%, SD = 21.1%; W =

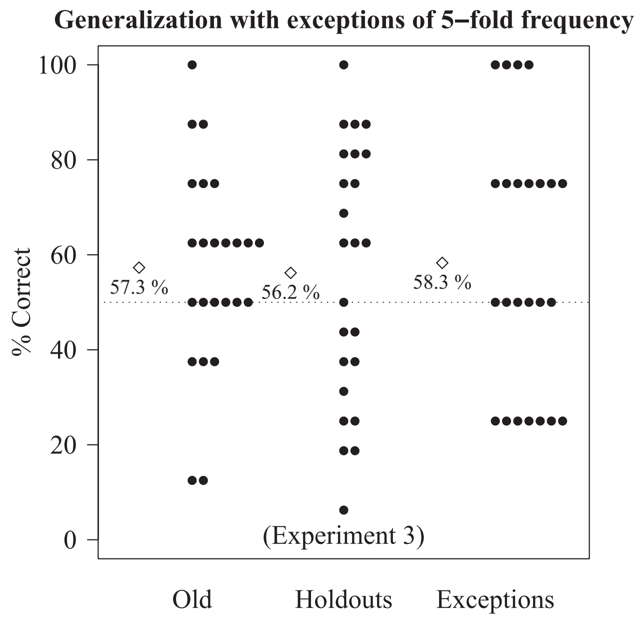

Figure 3. Results of Experiment 3. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When the token frequency of the exceptions was increased fivefold, participants failed to learn the majority pattern, but did not show evidence of learning the exceptions, either.

123.5, p = .095, Cohen's d = 0.34, 95% CI [48.36%, 66.22%], *ns*), holdout trials (M = 56.25%, SD = 26.9%; W = 177.5, p = .23, Cohen's d = 0.23, 95% CI [44.89%, 67.61%], *ns*), and exception trials (M = 58.33%, SD = 27.3%; W = 118.5, p = .13, Cohen's d = 0.31, 95% CI [46.83%, 69.84%], *ns*). An ANOVA on ranks with language (prefixation vs. suffixation default) as between-subjects factor and trial type (old vs. holdout vs. exception) as within-subjects factor revealed no main effect or interaction.

Performance in Experiment 3 was worse than in Experiment 2 on both old trials (U = 438, p = .002; permutation test: p = .0005) and holdout trials (U = 409, p = .013; permutation test: p = .0008). In contrast, performance on exception trials was better in Experiment 3 than in Experiment 2 (U = 174.5, p = .017; permutation test: p = .0105).

Experiment 4: Learning an Affixation Pattern With a High Frequency of Exceptions

Method

Experiment 4 was identical to Experiment 3 except that the exceptions were presented 42 times (as opposed to 15 times in Experiment 3). Their frequency was thus multiplied by 14 relative to Experiment 2 (where regular items and exceptions were presented three times).

We tested 16 new native speakers of American English (10 women, six men; mean age = 21.9 years; range: 13–34). Roughly the same number of participants started the experiment but were excluded from analysis because of computer problems at various points during familiarization or test.

Results

As shown in Figure 4, the participants' performance did not differ from chance in the old trials (M = 60.16%, SD = 27.1%; W = 94.5, p = .17, Cohen's d = 0.37, 95% CI [45.72%, 74.59%], *ns*) and in the holdout trials (M = 53.12%, SD = 21.8%; W = 60.5, p = .635, Cohen's d = 0.14, 95% CI [41.52%, 64.73%], *ns*). In contrast, performance on exception trials was reliably above chance (M = 71.88%, SD = 20.2%; W = 55, p = .005, Cohen's d = 1.1, 95% CI [61.13%, 82.62%]), with no difference between the language conditions (U = 21, p = .24, *ns*).

An ANOVA on ranks with language (prefixation vs. suffixation default) as a between-subjects factor and trial type (old vs. holdout vs. exception) as a within-subjects factor revealed no main effect or interaction.

There was no difference in performance in Experiment 4 (where exceptions were presented with 14-fold frequency) relative to Experiment 3 (where exceptions were presented with fivefold frequency) for either old trials (U = 219.5, p = .449; permutation test: p = .702, *ns*), holdout trials (U = 175, p = .647; permutation test: p = .692, *ns*), or exception trials (U = 247, p = .118; permutation test: p = .136, *ns*). Compared with that in Experiment 2 (where each exception and regular had the same frequency), the performance in Experiment 3 differed on old trials (U = 111.5, p = .025; permutation test: p = .012), holdout trials (U = 94.5, p = .007; permutation test: p = .0045) and exception trials (U = 311, p = .0008; permutation test: p = .0015).

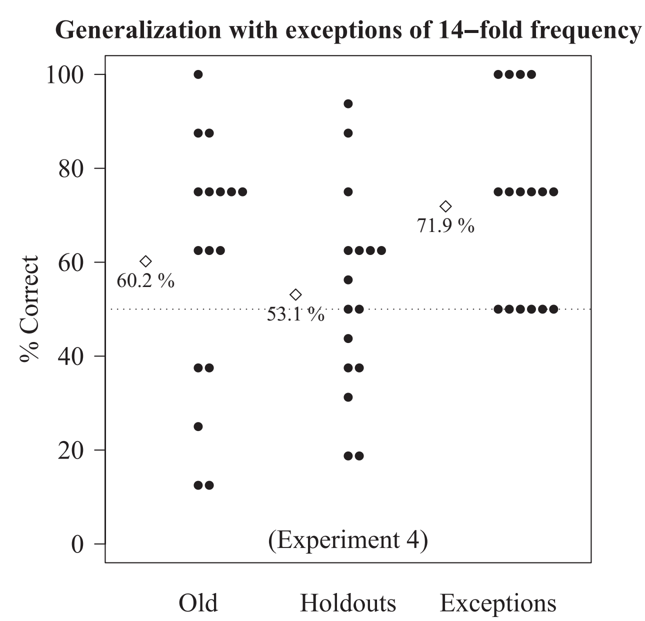

Figure 4. Results of Experiment 4. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When the token frequency of the exceptions was increased 14-fold, participants learned the exceptions, but showed no evidence of learning the overall pattern.

Figure 5 shows the combined results of Experiments 2–4. Our first analysis focused on the performance during holdout and exception trials, as these represent, respectively, generalization and exception learning, the two processes of primary interest. Regressing the performance on holdout trials against the token frequency of the exceptions, we obtained a slope of −.42, $F(1, 62) = 8.9$, $p = .004$, adjusted $R^2 = .111$. That is, for every 10% that the token frequency of the exception increased, the performance on holdout trials decreased by 4.2%. The regression of the performance on the exception trials against the token frequency of the exceptions yielded a slope of .654, $F(1, 62) = 16.95$, $p = .0001$, adjusted $R^2 = .202$. In other words, for every 10% that the token frequency of the exception increased, the performance on exception trials increased by 6.54%.

These results were confirmed by a mixed-effects model with the following fixed-effect predictors: trial type, language and ln(token frequency of the exceptions), Trial Type × ln(exception token frequency), Trial Type × Regularity. The only significant random effect was a slope adjustment to participants. We observed a significant positive slope adjustment for exception trials compared with holdout trials ($Z = 4.38$, $p < .0001$), whereas performance on holdout trials did not differ from that on old trials. Performance globally worsened with an increase in the logarithm of the type frequency of exceptions ($Z = 4.02$, $p < .0001$). Crucially, however, for exception trials, there was a large positive slope adjustment relative to the logarithm of the token frequency of the exceptions ($Z = 7.33$, $p < .0001$; see Table S.3.2 in the supplementary online material for more detailed results).

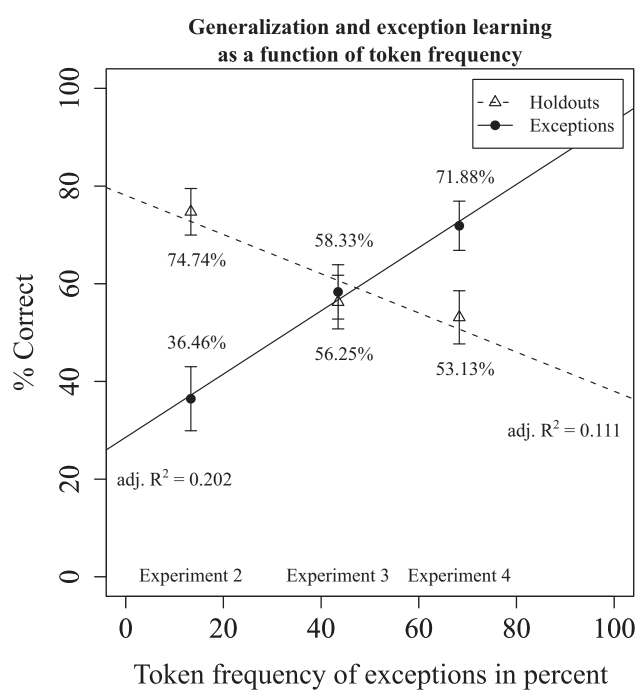

Figure 5. Results of Experiments 2–4. Triangles show the sample averages on holdout trials in the different experiments, circles the sample averages on exception trials, and error bars the sample standard errors of the means. Lines are regression lines. While the performance on exception trials increases as the token frequency of the exceptions increases, the performance on holdout trials decreases.

Discussion

In Experiments 2–4, participants were exposed to an affixation pattern in which the token frequency of the exceptions was varied. In all experiments, four out of the 30 familiarization items were exceptions; the type frequency of the exceptions was thus kept constant at 13.3%. The token frequency, in contrast, was manipulated such that 13.3% of the familiarization items were exceptions in Experiment 2 and 68.3% in Experiment 4. Results showed that the performance on exception trials increased with the token frequency of the exceptions. In contrast, the generalization performance on holdout trials decreased as the token frequency of the exceptions increased. Experiments 2–4 thus reveal two outcomes: When participants generalize the overall, majority pattern (as measured by their performance on holdout trials), they fail to learn the exceptions (Experiment 2); conversely, when they learn the exceptions (as in Experiment 4), they fail to learn the overall rule. These two outcomes are clearly not what children end up acquiring in natural languages. In the following experiments, we thus attempted to provide participants with additional cues that might allow them to learn both the overall regularity and the exceptions.

One such cue is the type frequency of the exceptions, that is, the number of different words that conform to the overall pattern and to the exception pattern. Although the type frequency of the overall pattern is very high in Experiments 2–4 (86.7% of the words are regular), it is possible, given the impoverished input of the test environment, that participants require a still higher type frequency of the regulars in order to learn the overall rule. We address this possibility in Experiment 5, where we halved the number of exceptions relative to Experiment 4 but presented each exception twice as often. As a consequence, the cumulative token frequency of the exceptions remained constant, whereas their type frequency decreased. If the type frequency of an inflectional pattern contributes to its productivity, participants may learn both the overall pattern and the exceptions under these conditions.

Experiment 5: The Role of Type Frequency

In Experiment 5, we kept the token frequency of the exceptions constant but decreased their type frequency.

Method

Experiment 5 was similar to Experiment 4 except that we used only two (instead of four) exceptions. The exception stems were /gA/ and /kweɪb/. Both stems were presented twice as often as the exception stems in Experiment 4. During test, we kept the number of exception trials constant by presenting each exception twice (rather than once as in Experiment 4). We tested 24 new native speakers of American English (15 women, nine men; mean age = 21.3 years; range: 15–38).

Results

The results of Experiment 5 are shown in Figure 6. In contrast to the previous experiments, participants performed above chance on all three trial types. They succeeded on old trials ($M = 70.83\%$, $SD = 26.8\%$; $W = 167.5$, $p = .003$, Cohen's $d = 0.78$, 95% CI [59.54%, 82.13%]), with no difference between the language conditions ($U = 67$, $p = .79$, *ns*). They also performed above chance

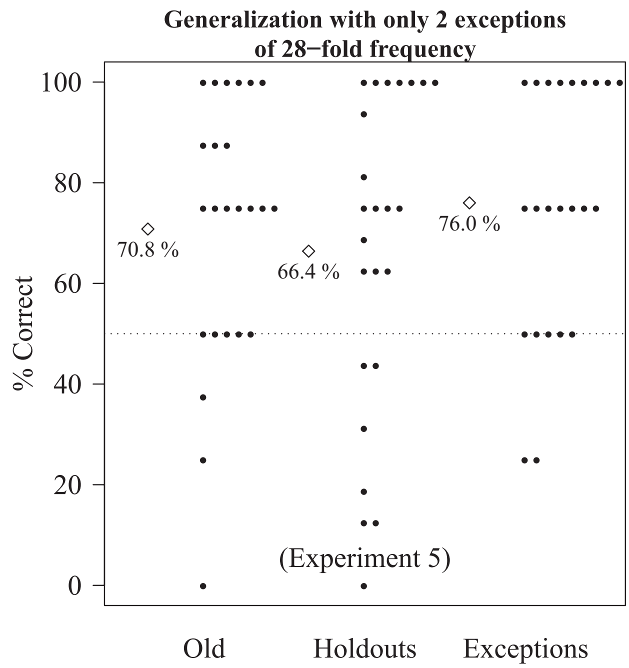

Figure 6. Results of Experiment 5. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When the type frequency of the exceptions was reduced, participants learned both the overall pattern and the exceptions even when 66.7% of the tokens were exceptions.

on holdout trials (M = 66.4%, SD = 32.3%; W = 228.5, p = .025, Cohen's d = 0.51, 95% CI [52.77%, 80.04%]), again with no difference between the language conditions (U = 67, p = .79, *ns*). Finally, they also succeeded on exception trials (M = 76.04%, SD = 25.0%; W = 180, p = .0005, Cohen's d = 1.0, 95% CI [65.5%, 86.59%]), again with no difference between the language conditions (U = 81, p = .60, *ns*). An ANOVA on ranks with language (prefixation vs. suffixation default) as a between-subjects factor and trial type (old vs. holdout vs. exception) as a within-subjects factor revealed no main effect or interaction.

Figure 7 contrasts the results of Experiments 4 and 5. Whereas the participants' performance on old trials did not differ between these experiments (U = 144, p = .182; permutation test: p = .204, *ns*) and exception trials (U = 166, p = .459; permutation test: p = .522, *ns*), their performance tended to be better on holdout trials in Experiment 5 than in Experiment 4, $U(N = 40) = 128$, p = .078 (permutation test: p = .118). As is clear from Figure 6, there was one outlier in Experiment 5 who differed by 2.65 standard deviations from the mean on old trials and by 2.06 standard deviations from the mean on holdout trials. Removing this participant yielded a significant difference between holdout trials in Experiments 4 and 5 (U = 256, p = .040; permutation test: p = .041), but not between old trials in these experiments (U = 240, p = .106; permutation test: p = .061, *ns*) or exception trials (U = 204, p = .558; permutation test: p = .602, *ns*).

Because the difference between holdout trials in Experiments 4 and 5 was marginal, we replicated these experiments with 48 new participants (27 women, 21 men; mean age = 21.7 years; range: 18–33) randomly assigned to the conditions corresponding to Experiments 4 and 5, respectively. In the replication of Experiment 4, participants failed on old trials (M = 55.73%, SD = 25.3%; W = 94.5, p = .174, Cohen's d = 0.23, 95% CI [45.06%, 66.4%], *ns*) and holdout trials (M = 53.65%, SD = 26.50%; W = 60.5, p = .635, Cohen's d = 0.14, 95% CI [42.45%, 64.84%], *ns*), but succeeded on exception trials (M = 71.88%, SD = 23.7%; W = 55.0, p = .005, Cohen's d = 0.92, 95% CI [61.88%, 81.87%]). An ANOVA on ranks with language (prefixation vs. suffixation default) as between-subjects factor and trial type as within-subjects factor revealed a main effect of trial type, $F(2, 44) = 6.4$, p = .0036, $p_{\text{perm}} = .0036$, $\eta_p^2 = .196$, and an interaction between trial type and language, $F(2, 44) = 4.26$, p = .02, $p_{\text{perm}} = .02$, $\eta_p^2 = .131$ (we discuss this latter interaction below).

In the replication of Experiment 5, in contrast, participants succeeded on old trials (M = 74.48%, SD = 24.0%; W = 231, $p < .001$, Cohen's d = 1.0, 95% CI [64.33%, 84.62%]), holdout trials (M = 72.66%, SD = 26.8%; W = 235, p = .003, Cohen's d = 0.84, 95% CI [61.33%, 83.98%]), and exception trials (M = 70.83%, SD = 29.2%; W = 194.5, $p < .005$, Cohen's d = 0.71, 95% CI [58.51%, 83.15%]). An ANOVA on ranks with language (prefixation vs. suffixation default) as a between-subjects factor and trial type as a within-subjects factor revealed no main effect or interaction. Figure 8 contrasts the results of the replications of Experiments 4 and 5. Whereas the participants' performance on exception trials did not differ between these experiments (U = 280, p = .87; permutation test: $p > .999$, *ns*), their performance was better in the replication of Experiment 5 than in the replication of Experiment 4 on old trials (U = 154, p = .005; permutation test: p = .023) and holdout trials (U = 151.5, $p < .005$; permutation test: p = .011). We thus replicated the main findings of Experiments 4 and 5 (except that the performance during old trials in the

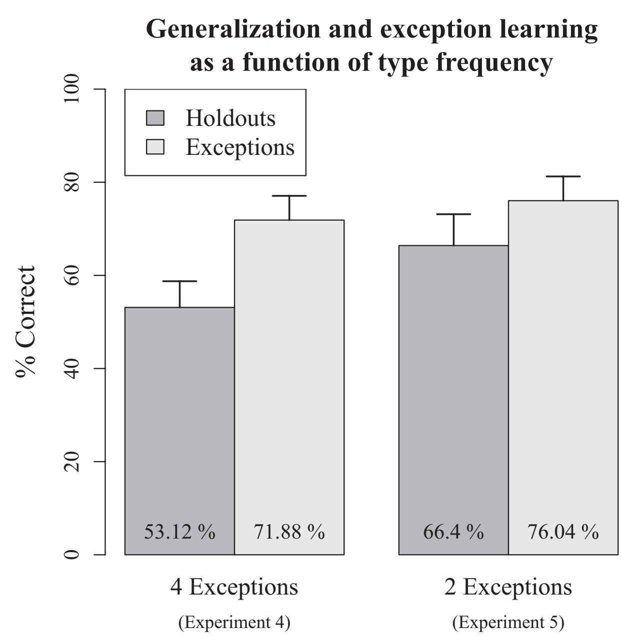

Figure 7. Results of Experiments 4 and 5. Bars show the sample averages and errors bars the sample standard errors. When the number of exceptions was reduced to two, participants learned both the exceptions and the majority pattern even when two thirds of the tokens were exceptions.

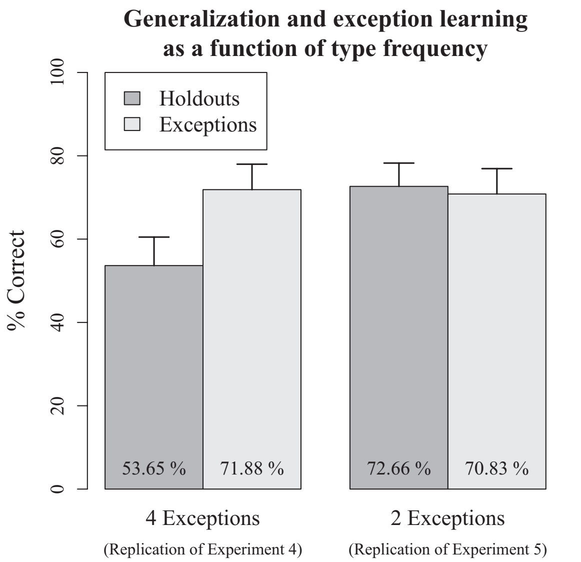

Figure 8. Results of the replications of Experiments 4 and 5. Bars show the sample averages and errors bars the sample standard errors. When the number of exceptions was reduced to two, participants learned both the exceptions and the majority pattern even when two thirds of the tokens were exceptions.

replication of Experiment 4 was somewhat lower than in the original Experiment 4).

These results are confirmed by a mixed-effects model with the following fixed-effect predictors: trial type, experiment (i.e., Experiment 4 and its replication vs. Experiment 5 and its replication), language, Trial Type × Experiment, Trial Type × Language. The only significant random effect was a slope adjustment for participants.

If decreasing the type frequency of the exceptions improves performance on holdout (and old) trials, two predictions follow. First, performance might improve in all trial types. On old and holdout trials, performance might improve because the type frequency of the exceptions is decreased. If the performance on exception trials in Experiment 4 was not at ceiling, it might improve in Experiment 5 as well, because each exception is presented twice as often compared with Experiment 4. Alternatively, we might also expect an interaction between experiment and trial type, as the effect of type frequency should be smaller for exception trials than for the other trial types.

Compared with holdout trials, performance was better on exception trials ($Z = 5.84$, $p < .0001$), whereas holdout trials did not differ from old trials (see Table S.3.3 in the supplementary online material for more detailed results). The performance in Experiment 5 (and its replication) was better than in Experiment 4 (and its replication; $Z = 4.05$, $p < .0001$). Compared with holdout trials, the effect of experiment was significantly less strong in exception trials ($Z = -2.94$, $p = .0033$). (This corresponds to an interaction between experiment and trial type in conventional ANOVA analyses.)

Finally, participants performed significantly better in the prefixation condition compared with the suffixation condition ($Z = 3.42$, $p = .001$), an effect that was significantly stronger in holdout trials than in exception trials ($Z = 1.15$, $p < .0001$).

Discussion

In contrast to Experiments 2–4, participants in Experiment 5 succeeded on all three trial types. They thus learned the overall, majority pattern and generalized it in holdout trials, while simultaneously learning the exceptions. Their successful acquisition of the majority pattern is remarkable given that 66.7% of the familiarization items were exceptions. Participants were thus twice as likely to hear an exception as a regular item. As Experiments 2–4 showed that a high token frequency of the exceptions negatively affected the learning of the default pattern, one might have expected a failure to learn the default pattern. Nonetheless, participants learned the majority pattern in Experiment 5.

We believe that the most likely explanation for the successful acquisition of the majority pattern is that Experiment 5 used only half the number of exceptions as in Experiment 4. Hence, the type frequency of the exceptions was halved, and that of the regular default pattern increased. Apparently, a very high type frequency provides a kind of immunity to the regular pattern, protecting it from being overwhelmed by exceptions of high token frequency.

Before accepting this conclusion, it is necessary to rule out an alternative interpretation of the difference between Experiments 4 and 5. Indeed, compared with Experiment 4, Experiment 5 introduced two changes. On the one hand, and as mentioned above, the type frequency of the exceptions was halved. On the other hand, by keeping the token frequency of the exceptions constant in Experiments 4 and 5, the exceptions in Experiment 5 were much more frequent relative to the regulars than in Experiment 4; in Experiment 5, each exception occurred 28 times as often as each regular item, whereas in Experiment 4, each exception occurred only 14 times as often as each regular item. The difference in relative frequency might have protected the regular pattern from being overwhelmed by token frequency in a different way: In line with previous suggestions (Bybee, 1995; see also Bailey & Hahn, 2001, for similar findings in the domain of wordlikeness), participants might ignore extremely frequent items, with the consequence that the exceptions in Experiment 5 would not enter the computations that allow participants to generalize the regular pattern. If extremely frequent items are ignored, they cannot interfere with the generalizations, either.

Experiment 6 was designed to rule out the possibility that the successful generalization of the regular pattern in Experiment 5 was exclusively due to the extremely high frequency of the exceptions relative to the regulars. Experiment 6 was identical to Experiment 4 except that each exception was presented twice as frequently—that is, 28 rather than 14 times. Compared with that of regular items, therefore, the relative frequency of exceptions is the same in Experiments 5 and 6; however, as Experiment 6 used four exceptions, the type frequency of the exceptions was the same as in Experiment 4 and thus twice that used in Experiment 5.

Experiment 6: The Role of the Relative Frequency of Exceptions and Regular Items

Method

Experiment 6 was similar to Experiment 4 except that we played each exception 28 times rather than 14 times. We tested 20 new native speakers of American English (13 women, 7 men; mean age = 22.1 years; range: 18–35).

Results

The results of Experiment 6 are shown in Figure 9. As in Experiment 4, the participants' performance differed from chance neither on the old trials (M = 56.25%, SD = 32.57%; W = 82.5, p = .464, Cohen's d = 0.19, 95% CI [41.01%, 71.50%], *ns*) nor on the holdout trials (M = 47.5%, SD = 27.4%; W = 86.5, p = .747, Cohen's d = 0.09, 95% CI [34.68%, 60.32%], *ns*). In contrast, the performance on exception trials was reliably above chance (M = 67.5%, SD = 25.8%; W = 103.5, p = .011, Cohen's d = 0.68, 95% CI [55.44%, 79.56%]). An ANOVA on ranks with language (prefixation vs. suffixation default) as between-subjects factor and trial type (old vs. holdout vs. exception) as within-subjects factor revealed a marginally significant main effect of trial type, $F(2, 36) = 2.98$, $p < .063$, $p_{\text{perm}} < .064$, $\eta_p^2 = .142$, but no other main effects or interactions.

In a comparison of Experiments 4 and 6, there was no difference in performance on either old trials (U = 169, p = .784; permutation test: p = .788, *ns*), holdout trials (U = 176, p = .621; permutation test: p = .410, *ns*), or exception trials (U = 172, p = .701; permutation test: p = .661, *ns*). In contrast, participants performed better on holdout trials in Experiment 5 than in Experiment 6, $U(N = 36) = 332$, p = .03 (permutation test: p = .022), but the participants' performance differed neither on old trials (U = 301.5, p = .143; permutation test: p = .146, *ns*) nor on exception trials (U = 286, p = .262; permutation test: p = .252, *ns*).

A comparison of Experiment 6 and the replications of Experiments 4 and 5 yielded similar results. In a comparison of Experiment 6 and the replication of Experiment 4, the participants' performance differed neither on old trials (U = 228.5, p = .793; permutation test: p = .839, *ns*) nor on holdout trials (U = 273, p = .442; permutation test: p = .692, *ns*) nor on exception trials (U = 264.5, p = .551; permutation test: p = .717, *ns*). In contrast, participants performed better on holdout trials in the replication of Experiment 5 than in Experiment 6 (U = 369, p = .002; permutation test: p = .008), but the participants' performance differed neither on old trials (U = 317.5, p = .066; permutation test: p = .057, *ns*) nor on exception trials (U = 264.5, p = .555; permutation test: p = .772, *ns*).

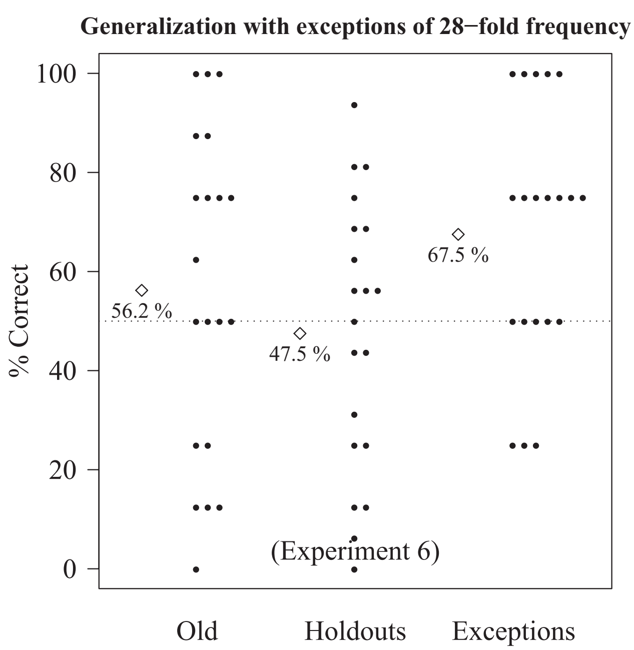

Figure 9. Results of Experiment 6. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When the exceptions were played 28 times as often as the regular items, participants learned only the exceptions but not the overall pattern.

We confirmed these analyses using a mixed-effects model with the following fixed-effect predictors: trial type, experiment (combining Experiments 4 and 5 with their respective replications), language. The only significant random effect was a slope adjustment for participants.

In a comparison of Experiment 6 to Experiment 4 and its replication, performance on exception trials was significantly better than on holdout trials (Z = 4.49, $p < .0001$), whereas there was no difference between holdout trials and old trials (see Table S.3.4 in the supplementary online material for more detailed results). Further, participants in the suffixation group performed significantly better than participants in the prefixation group (Z = 3.03, p = .002).

In a comparison of Experiment 6 to Experiment 5 and its replication, performance on exception trials was significantly better than on holdout trials (Z = 2.10, $p < .036$), whereas there was no difference between holdout trials and old trials (see Table S.3.5 in the supplementary online material for more detailed results). Crucially, performance in Experiment 6 was significantly worse than performance in Experiment 5 and its replication ($Z = -3.38$, p = .0007). Compared with that of holdout trials, performance on exception trials deteriorated significantly less between Experiments 5 and 6 (Z = 2.17, p = .030). Finally, participants in the suffixation group performed significantly better than participants in the prefixation group (Z = 2.0, p = .0465), an effect that was marginally stronger for holdout trials compared with exception trials (Z = 1.91, p = .056).

Discussion

The results of Experiment 6 replicate those of Experiment 4. When the exceptions were extremely frequent relative to the regulars, without reducing the type frequency of the exceptions, participants learned only the exceptions but not the overall pattern. This suggests that the successful learning of both the overall pattern and the exceptions in Experiment 5 was due at least in part to the reduction in type frequency of the exceptions.

The reduction in type frequency of the exceptions in Experiment 5 might have influenced the learning of the overall pattern in two different ways. On the one hand, decreasing the type frequency of the exceptions increases the type frequency of the majority pattern, which may make it easier to learn. On the other hand, halving the type frequency of the exceptions may make the exception pattern difficult to learn, which may reduce its interference with the majority pattern. In fact, as the performance on holdout trials decreased markedly between Experiments 1 and 2 (where the only difference was the presence of four exceptions in Experiment 2 and no exceptions in Experiment 1), it is plausible that the mere presence of a rare second affixation pattern can interfere with and negatively affect the learning of the majority pattern.

Although our experiments do not allow us to distinguish between these two hypotheses, we note that the relative change in type frequency is much larger for the exceptions (i.e., a change of 6.7% relative to a type frequency of 13.3% in Experiment 4) than for the overall pattern (i.e., a change of 6.7% relative to a type frequency of 86.7% in Experiment 4). Hence, to the extent that the pattern of the exceptions is learned, one would expect the change in type frequency between Experiments 4 and 5 to have a much larger influence on the learning of the exception pattern than on that of the majority pattern. In either case, however, the combined results of Experiments 4–6 demonstrate that the type frequency of an affixation pattern strongly influences how well it can be learned despite massive interference from alternative patterns with high token frequencies.

Experiment 7: Relative or Absolute Token Frequencies? Part 1

Experiments 1–4 show that a high token frequency of exceptions can overwhelm the learning of an affixation pattern, whereas Experiments 4–6 show that a very high type frequency protects the overall pattern from being overwhelmed by exceptions with high token frequencies. However, Experiments 1 and 4 left unclear the precise role of the token frequency. As mentioned above, items with high token frequencies have high absolute numbers of occurrences and occur often relative to other items with lower token frequencies. Although it is generally assumed that high token frequencies help learning exceptions because the exceptions occur (absolutely) more often if they have high token frequencies (e.g., Bybee, 1995; Marcus et al., 1995; Pinker, 1991; Plunkett & Marchman, 1993; Tomasello, 2000), this assumption has never been investigated empirically. We address this issue in Experiments 7 and 8. In these experiments, we used a design similar to the one implemented in Experiment 5, where the exceptions are sufficiently frequent to be learned but their type frequency is sufficiently low for the overall, majority pattern to be learned. We then contrasted this familiarization with one in which the exception occurs as often (in absolute terms) as in the high frequency situation but where all words have the same (relative) frequency.

The purpose of Experiment 7 was, therefore, to establish another context in which both the overall pattern and the exceptions are learned without resorting to excessively long familiarizations. Experiment 7 was thus identical to Experiment 5, except that the exceptions were only 7 times as frequent as the regular items (as opposed to 28 times as frequent in Experiment 5).

Method

Experiment 7 was identical to Experiment 5 except for the frequency of the exceptions. Rather than presenting them 28 times as often as the regular items, we increased their frequency only sevenfold. We tested 20 new native speakers of American English (13 women, seven men; mean age = 21.1 years; range: 18–32).

Results and Discussion

As shown in Figure 10, participants performed above chance in the old trials (M = 73.12%, SD = 31.2%; W = 130, p = .011, Cohen's d = 0.74, 95% CI [58.51%, 87.74%]), holdout trials (M = 67.19%, SD = 25.9%; W = 116.5, p = .013, Cohen's d = 0.66, 95% CI [55.08%, 79.3%]), and exception trials (M = 67.5%, SD = 29.4%; W = 99, p = .024, Cohen's d = 0.6, 95% CI [53.76%, 81.24%]). An ANOVA on ranks with language (prefixation vs. suffixation default) as between-subjects factor and trial type (old vs. holdout vs. exception) as within-subjects factor revealed no main effect or interaction.

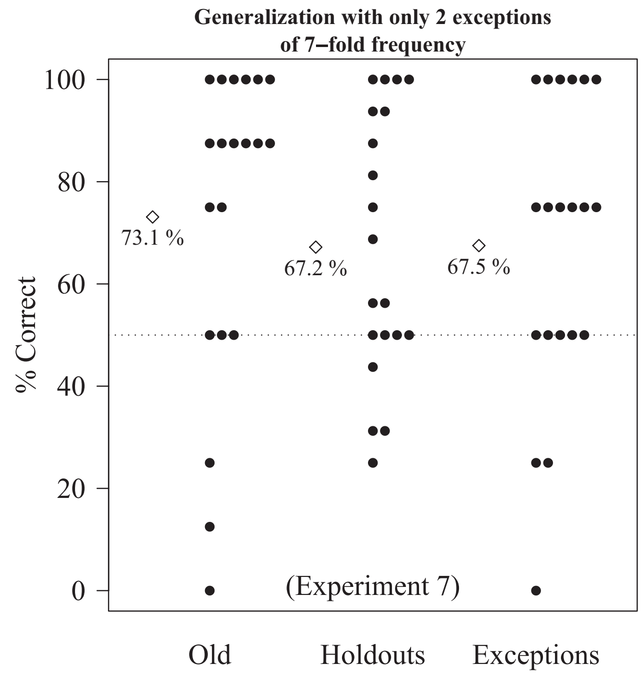

Figure 10. Results of Experiment 7. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When two out of 30 items were exceptions, and when the exceptions were presented 7 times as frequently as the regular items, participants learned both the majority pattern and the exceptions.

Experiment 8: Relative or Absolute Token Frequencies? Part 2

In Experiment 8, each exception occurred as often as in Experiment 7, thus equating the absolute number of occurrences. However, the number of occurrences of the regular items was increased such that each regular item occurred as often as each exception. In other words, although the (absolute) number of occurrences of the exceptions was the same in Experiments 7 and 8, in Experiment 8, the exceptions were no longer relatively more frequent than the regular items. If relative frequency has an effect on the acquisition of exceptions over and above the absolute number of occurrences, then we predict a difference in the performance on exception trials between Experiments 7 and 8.

Method

We tested 20 new native speakers of American English (15 women, five men; mean age 23.6 years; range: 15–30). The familiarization phase was similar to that in Experiment 7 except that all words now occurred equally often. Whereas regular words occurred three times in the familiarization of Experiment 7 and the two exceptions 21 times, all words occurred 21 times in the familiarization phase of Experiment 8. The test phase was identical to that in Experiment 7.

Results

As shown in Figure 11, participants performed above chance on old trials (M = 76.88%, SD = 25.4%; W =143.5, p = .0015, Cohen's d = 1.1, 95% CI [64.98%, 88.77%]), holdout trials (M = 85.0%, SD = 22.8%; W = 202.5, p = .0003, Cohen's d = 1.5, 95% CI [74.33%, 95.67%]), and exception trials (M = 63.75%, SD = 35.8%; W = 90, p = .082, Cohen's d = 0.38, 95% CI [47.0%, 80.5%]). An ANOVA on ranks with language (prefixation vs. suffixation default) as a between-subjects factor and trial type (old vs. holdout vs. exception) as a within-subjects factor revealed no main effect or interaction.

As shown in Figure 12, the performance on exception trials did not differ between Experiments 7 and 8 (U = 207.5, p = .845; permutation test: p = .847, *ns*). Whereas the participants' performance on old trials did not differ between these experiments (U = .192, p = .835; permutation test: p = .664, *ns*), their performance was better on holdout trials in Experiment 8 than in Experiment 7 (U = 119.5, p = .027; permutation test: p = .026).

We compared the results of Experiments 7 and 8 using a mixed-effects model with the following fixed-effect predictors: trial type, experiment, language, Trial Type × Experiment, Trial Type × Language. The only significant random effect was a slope adjustment for participants (see Table S.3.6 in the supplementary online material for more detailed results). Overall, participants performed better in Experiment 8 compared with Experiment 7 (Z = 2.64, p = .011). This improvement was significantly stronger for holdout trials compared with either old trials (Z = 2.7, p = .007) or exception trials (Z = 3.36, p = .0008).

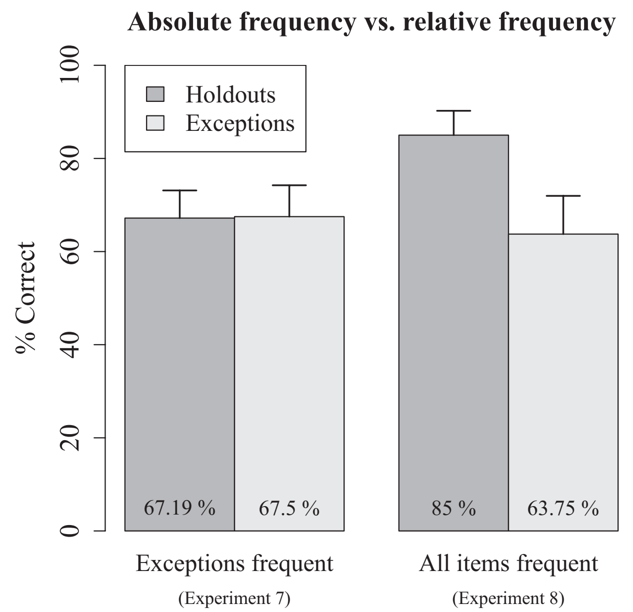

Figure 12. Results of Experiments 7 and 8. Bars show the sample averages and errors bars the sample standard errors. In both experiments, exceptions occurred equally often, but only in Experiment 7 were exceptions frequent relative to regulars. Results showed no difference in the performance on exception trials between the two experiments.

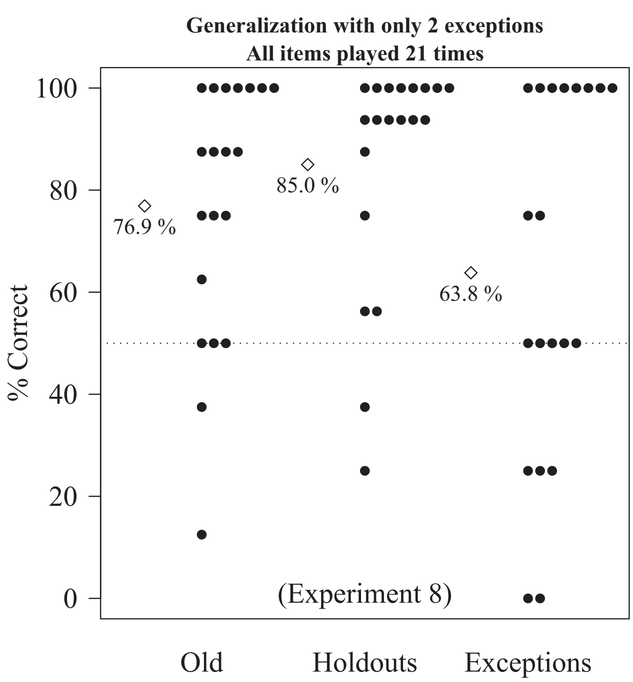

Figure 11. Results of Experiment 8. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When two out of 30 items were exceptions, and when all items were presented equally frequently, but very often, participants learned both the majority pattern and the exceptions.

Discussion

There seems to be a consensus that a high token frequency of exceptions facilitates the process of language acquisition (e.g., Bybee, 1995; Marcus et al., 1995; Pinker, 1991; Plunkett & Marchman, 1993; Tomasello, 2000), a conclusion that is also licensed by the results of Experiments 2–4, where participants performed better on exception trials as the token frequency of the exceptions increased. However, a high token frequency could further the acquisition of exceptions in two ways. On the one hand, high-frequency items are repeated often; in line with Ebbinghaus's (1885/1913) observation that repeated items are memorized better, this may allow learners to acquire exceptions better. On the other hand, high-frequency items also occur more often relative to lower frequency items; this may make the high frequency items stand out and facilitate their acquisition in turn.

Experiments 7 and 8 provide an initial step in addressing this issue. In these experiments, the absolute number of occurrences of the exceptions was matched. However, in Experiment 8, regular items occurred as often as the exceptions, whereas exceptions were relatively more frequent in Experiment 7. As we did not observe any difference in the participants' performance on the exception trials between these experiments, we conclude that token frequency does not contribute to the acquisition of exceptions over and above the absolute number of occurrences of the exceptions and that exceptions do not need to be relatively more frequent than regulars to be learned. In line with Ebbinghaus's (1885/1913) results, therefore, they just need to occur sufficiently often.

Overall Analysis

To analyze our data further, we performed a combined analysis of Experiments 2–8 (i.e., all experiments with exceptions). Even though this data set is extremely unbalanced (e.g., all high frequency items are exceptions), such an analysis might provide additional confirmatory evidence for the results of the more "local" analyses performed above. As a first step, we fitted a generalized linear model to our results with the empirical logistic function as a link function (i.e., the functions used to transform the data) and the following fixed-effect predictors: trial type, type frequency, token frequency, language (prefixation vs. suffixation), Trial Type × Type Frequency, Trial Type × Token Frequency, Trial Type × Language. Further, we included random intercepts for participants, stems, and the trial number. After comparing the likelihoods of the models (using likelihood ratio tests), we kept all the fixed factors but only the random intercepts for participants.

The results of this analysis are shown in Table 2. As main effects of test type are not meaningful, we focused on the effects of type and token frequency and their interactions with test type. Both type and token frequency showed a strong positive correlation with performance ($Z = 4.16$, $p < .0001$, and $Z = 4.97$, $p < .0001$, respectively). Crucially, performance on holdout trials showed a stronger influence of the type frequency than the performance on exception trials ($Z = 4.28$, $p < .0001$); in contrast, the influence of the type frequency did not differ between holdout trials and old trials ($Z = 1.31$, $p = .518$, *ns*). There was no interaction between token frequency and test item type, either. Finally, participants in the various suffixation groups performed better than those in the prefixation groups ($Z = 3.4$, $p = .0007$), an effect that was significantly stronger in holdout trials compared with exception trials ($Z = 2.31$, $p = .021$); in contrast, the effect of language did not differ between holdout trials and old trials. The latter interaction has a straightforward interpretation: If suffixation patterns are easier to learn than prefixation patterns, the advantage for such patterns should be observed in old and holdout trials rather than in exception trials, simply because the correct choices in exception trials deviate from the pattern that is easier to learn.

Together, these results allow for three conclusions. First, the type frequency of a pattern plays an important role in assuring its acquisition; the effect of type frequency was much larger on old and holdout trials compared with exception trials, suggesting that it allowed participants to generalize it as the default pattern. Second, the token frequency seems to allow participants to memorize regular forms and exceptions alike. Third, our finding that suffixation patterns are learned better than prefixation patterns (see also Creel & Dahan, 2010, for related results) might relate to the cross-linguistic observation that suffixation is much more frequent than prefixation (e.g., Dryer, 2005). Possibly, it is more frequent because it is easier to learn, making it more stable in the course of language change. Alternatively, given that English uses only suffixes but no prefixes for morphological purposes, and that prefixes are restricted to "semantic" roles (e.g., *re-*, *un-*), our participants might have performed better on suffixes for this reason as well.

General Discussion

For over 2 decades now, studies of inflectional morphology have been at the center of important debates about the fundamental nature of cognitive processes and, of course, the structure of language more specifically (e.g., Butterworth, 1983; Bybee, 1995; Caramazza et al., 1988, 1985; Clahsen et al., 1992; Dabrowska, 2001; Hahn & Nakisa, 2000; Hare & Elman, 1995; Hare et al., 1995; Joanisse & Seidenberg, 1999; Kiparsky, 1982; Marcus et al., 1995; Marslen-Wilson & Tyler, 1997, 1998; Pinker, 1991, 1999; Pinker & Prince, 1988; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986; Seidenberg, 1997; Taft, 1979; Taft & Forster, 1975; Tyler et al., 2005). Despite its importance, the factors that influence its acquisition have received little direct experimental scrutiny and are currently debated. One important controversy in this debate concerns the influence of type frequency on the productivity of inflectional patterns; to date, this problem has never been investigated in learning experiments. Here our aim was to provide one way into this problem by using an artificial grammar learning paradigm to explore the influence of type and token frequency on the acquisition of an overall affixation pattern (that applies to the

Table 2
Results of a Mixed-Effects Model Comparing the Results of Experiments 2–8

Predictor	Estimate	*SE*	*Z*	*p*
Intercept	2.2	0.29	7.45	<.0001
Test type				
Old	−0.22	0.26	−0.84	.399
Exception	**−3.55**	**0.92**	**−3.84**	**.00012**
ln(type frequency)	**9.69**	**2.33**	**4.16**	**<.0001**
ln(token frequency)	**0.8**	**0.16**	**4.97**	**<.0001**
Language				
Suffix	**0.57**	**0.17**	**3.4**	**.0007**
Test type = old × ln(type frequency)	−1.31	2.03	−0.65	.518
Test type = exception × ln(type frequency)	**−10.83**	**2.53**	**−4.28**	**<.0001**
Test type = old × ln(token frequency)	−0.24	0.15	−1.61	.108
Test type = exception × ln(token frequency)	−0.16	0.25	−0.64	.521
Test type = old × Language = suffix	0.07	0.15	0.46	.646
Test type = exception × Language = suffix	**−0.43**	**0.19**	**−2.31**	**.0208**

Note. Significant predictors are shown in bold. ln = natural logarithm.

greatest number of “stems”) and on the acquisition of exceptions to that pattern.

In Experiment 1, we showed that participants learned the affixation patterns almost perfectly when no exceptions were present. In Experiment 2, we showed that the mere presence of a few exceptions decreased the participants’ learning of the overall, majority pattern, even though participants did not seem to learn the exceptions; rather, they tended to overextend the majority pattern to the exceptions. Experiments 2–4 showed that participants learned the exceptions better as their token frequency increased but that their performance on regular items simultaneously decreased to the point that it did not differ from chance when they performed above chance on the exception trials. In Experiment 5, we attempted to allow participants to learn both the majority rule and the exceptions to that rule by manipulating the type frequency of the exceptions. When the type frequency of the exceptions was decreased, and that of the regular items increased, participants learned both the overall pattern and the exceptions to it—even when two thirds of the tokens were exceptions. Experiment 6 further showed that the successful acquisition of both the overall pattern and the exceptions in Experiment 5 was not simply due to the extremely high token frequency of the exceptions.

Finally, Experiments 7 and 8 showed that a high token frequency facilitates learning the exceptions because it ensures that the absolute number of occurrences of the exceptions is high, and not because it ensures that the exceptions are frequent relative to other items. This suggests that the main role of a high token frequency of the exceptions is to ensure that they occur sufficiently often to be memorized.

In summary, the results from these eight experiments suggest that type and token frequencies have different roles in the learning of exceptions and a regular pattern. The token frequency seems to ensure that exceptions are heard sufficiently often so that they can be memorized. That is, repeating items more often improves performance on these items irrespective of whether they are regular or exceptions; however, a high number of repetitions is necessary for the exceptions to be learned, and the regular patterns can be acquired with lower token frequencies as well. The type frequency, in contrast, seems to determine the productivity of a pattern.

Of course, there are important differences between our experiments and natural language acquisition. First, using an adult population to test questions of language acquisition always raises the question of whether the results obtained generalize in a significant way to infants. Given the scarcity of direct experimental evidence for the role of type and token frequency on the acquisition of affixation patterns, however, we just started investigating the respective roles of type and token frequency in adults, and leave for further study the question of whether infants would behave qualitatively similarly to adults.

Second, to our knowledge, the specific alternation we use—between prefixation and suffixation—does not exist in any natural language. However, the point of these experiments is not to simulate a specific language faithfully, but rather to use the simplest possible situation to investigate the role of type and token frequency for learning affixation-like rules.

Third, the number of items and the type and token frequency used in these experiments clearly do not match those in natural languages, nor does the uniform frequency distribution of words in our experiments mirror the Zipfian pattern typically found in natural languages. Moreover, the items used in our experiments have no meaning, and an important part of language acquisition is to acquire the meaning of words at the same time as their grammar. However, given that our participants had less than 20 min to learn our miniature language (and not several years as real children), we assume here that our results would scale up to natural language acquisition, but it is an important topic for further research to establish whether this assumption is licensed.

Type Frequency or Phonological Diversity?

The results of Experiments 4–6 suggest that increasing the type frequency of the overall pattern and decreasing that of the exceptions protects the overall pattern from being overwhelmed by the exceptions—even when two thirds of the tokens are exceptions. These results seem to suggest that the type frequency of a pattern determines its productivity. However, it is possible that our results may also be influenced by a closely related factor, namely the phonological diversity of the items carrying a pattern. Indeed, if many different items follow a pattern, it is likely that this pattern is also followed by a phonologically rather diverse set of items. There have been suggestions that not only the type frequency of a pattern influences its productivity but also the phonological diversity of the items following that pattern (e.g., Bybee, 1995; Bybee & Moder, 1983). This suggestion also receives support from the observation that learners consider the phonological properties of a word when inflecting it (e.g., Albright & Hayes, 2003; Prasada & Pinker, 1993); for example, verbs similar to *sing* tend to have a past tense that is similar to *sang*.

Although our experiments do not allow us to decide whether the difference between Experiments 4 and 5 was due to a change in the phonological diversity of the regular items or their type frequency, our familiarization items were designed to be as diverse as possible for monosyllabic items (see Table S.1 in the supplementary online material for the complete list of stems). It thus seems plausible to conclude that removing two regular stems does not dramatically influence the phonological diversity of the regulars and that the difference between Experiments 4 and 5 is more likely to be due to a change in type frequency.

What Is the Role of a High Type Frequency?

Although the results of Experiments 4 and 5 suggest an important role of type frequency for making an affixation pattern productive, it is unclear in which way the type frequency acts. In fact, there are at least two significant differences between Experiments 4 and 5. On the one hand, the type frequency of the overall pattern was increased from 86.67% to 93.33%; on the other hand, the type frequency of the exceptions was decreased from 13.33% to 6.67%.

These changes may have had two (related) consequences. First, the increase in the type frequency of the regulars may make the learning of the overall pattern more robust; if so, the type frequency of this pattern would determine its productivity. Alternatively, the decrease in the type frequency of the exceptions may make the exception pattern less productive, and thus reduce interference with the overall pattern. To see this point, it is important to note that our exceptions were not arbitrary, idiosyncratic items (such as *go–went*), but rather obeyed a different pattern than the majority one: When the majority pattern was prefixation, the exceptions were suffixed, and when the majority pattern was suffixation, the exceptions were prefixed. Hence, participants may have learned both the prefixation and the suffixation pattern as patterns, and both may have interfered with each other. Once the type frequency of the exceptions drops under a certain threshold, this pattern may no longer be learned and thus stops interfering with the overall pattern.

Although our results do not allow us to distinguish between these possibilities, both would demonstrate an important role of the type frequency of a pattern in its productivity, either that of the overall pattern or that of the exceptions. There are, however, two considerations that make the former possibility more plausible. First, as outlined in the previous section, phonological diversity of a pattern may influence how productive it is. Because both the overall and the exception rule were applied to items that were phonologically quite diverse, it is also possible that the exception pattern may have been learned, and may have interfered with the overall pattern. Second, the relative change of the type frequencies is much greater for the exceptions than for the overall pattern; whereas the type frequency of the exceptions was halved from Experiment 4 to Experiment 5, the type frequency of the exceptions increased only 7.7% relative to the type frequency of 86.67% in Experiment 4. Hence, one would expect this change to have a more significant impact on the learning of the exception pattern than that of the overall rule. If so, the better learning of the overall pattern in Experiment 5 compared with Experiment 4 would reflect a decrease in the interference between the two patterns. Independently of whether this interpretation holds up, however, the type frequency of a pattern seems to determine how productive it is.

Implications for Models of Language Acquisition

Although the controversy about the role of type and token frequencies for acquiring inflectional patterns has played a prominent role in past and ongoing debates about the symbolic versus statistical nature of inflectional morphology (e.g., Bybee, 1995; Marcus et al., 1995), we believe that the extremely simple rules used here allow us to make an equally simple point about the specific contributions of these two mechanisms in the acquisition process. Specifically, to learn whether a specific syllable occurs word-initially (in a prefixation pattern) or word-finally (in a suffixation pattern), one needs to represent the initial versus the final positions within the words. Representing such positions, however, is extremely difficult for statistical mechanisms such as standard connectionist networks (e.g., Endress & Bonatti, 2007; Endress & Mehler, 2009; see also Henson, 1998, for a similar argument in the context of short-term memory). To acquire the affixation patterns, learners thus need at least some minimal representational machinery, involving the ability to represent the first versus the last position within words, and this machinery is unlikely to arise from purely associative mechanisms (Endress, Nespor, & Mehler, 2009). Moreover, this ability is by no means specific to language or to humans, but rather shared with various other primate species (Endress, Cahill, et al., 2009; Endress, Carden, et al., 2010; Orlov, Yakovlev, Hochstein, & Zohary, 2000; Terrace, 2005) and, presumably, other animals.

Our results also demonstrate that learners are sensitive to distributional information such as the type frequency and the token frequency of the regular items and the exceptions. We thus suggest that both symbolic and statistical mechanisms contribute to learning inflectional patterns: Some rule-based machinery is necessary to represent the transformation involved in the patterns, but to actually acquire them, learners have to deploy distributional mechanisms to determine which patterns actually occur.

The interpretation we offer of our results is consistent with a broad range of models that incorporate both a sensitivity to distributional information and a representational system sufficiently rich to capture the patterns to be acquired, an approach that has been very successful in computational linguistics (e.g., Bod & Scha, 1997; Manning & Schütze, 1999), including inflectional morphology (e.g., Albright, 2002; Albright & Hayes, 2003). Models that do not incorporate these components, in contrast, are ruled out by our data.

In the context of such models, our results suggest that type and token frequency have different roles in the acquisition of inflectional patterns. Whereas the type frequency influences the productivity of a pattern, the token frequency guarantees that patterns with low type frequencies can be memorized. Of course, there may be other determinants of the productivity of a pattern, such as the phonological or semantic diversity of the items to which a pattern applies, and these factors deserve to be investigated experimentally. Further, it will be important to explore how results such as ours scale up to different rules, and different relationships between type and token frequency.

How Do Type and Token Frequencies Affect Learning and Memory?

Our results show that both token and type frequency influence how inflectional patterns are acquired. But how do they act? We believe that both kinds of frequency have straightforward psychological interpretations. The role of the token frequency is readily interpreted in terms of Ebbinghaus's (1885/1913) finding that repeated items are remembered better; accordingly, the representations of items that occur more frequently might simply be stronger.

There is an equally simple interpretation for the role of the type frequency. Possibly, children (and adults in our experiments) learn about inflectional patterns only when they encounter a word for the first time. For example, when they encounter a suffixed word for the first time, they might increase the strength of the representation of the suffixation pattern. When they subsequently encounter this word, they might not update the pattern's representation. When encountering a new suffixed word for the first time, in contrast, they would increase the representational strength of the corresponding pattern. As a result, the representation of the pattern would be strengthened only once for each word, namely when the learner encounters the word for the first time. Consequently, the representational strength of the pattern would depend only on the number of word types, irrespective of their frequency. (If learners need more than one encounter with a word in order to remember it, they might strengthen the representation of the corresponding patterns more than once per word; however, as the number of encounters needed to memorize a word is probably constant, the resulting representational strength would still depend on the number of word types, but not on their token frequency.)

In sum, both type frequency and token frequency might influence the acquisition of words and the corresponding patterns by virtue of psychological principles of representational strength. Even though these principles are unlikely to be specific to verbal material, and are certainly not sufficient for language acquisition, the language faculty might exploit them for morphological (and possibly other) processing. If correct, this would provide an important case study for how originally nonlinguistic processes interact with the human-specific computations of the language faculty.

References

Albright, A. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language, 78,* 684–709. doi:10.1353/lan.2003.0002

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

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Baayen, R. H., & Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language, 81,* 666–698. doi:10.1353/lan.2005.0112

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44,* 568–591. doi:10.1006/jmla.2000.2756

Balling, L. W., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes, 23,* 1159–1190. doi:10.1080/01690960802201010

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

Bod, R., & Scha, R. (1997). Data-oriented language processing. In S. Young & G. Bloothooft (Eds.), *Corpus-based methods in language and speech processing* (pp. 137–173). Boston, MA: Kluwer Academic.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production: Vol. 2. Development, writing and other language processes* (pp. 257–294). London, England: Academic Press.

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

Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. *Language, 59,* 251–270. doi:10.2307/413574

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297–332. doi:10.1016/0010-0277(88)90017-0

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology, 2,* 81–114. doi:10.1080/02643298508252862

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65. doi:10.1093/mind/os-XI.41.63

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition, 45,* 225–255. doi:10.1016/0010-0277(92)90018-D

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. doi:10.1037/0033-295X.108.1.204

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes, 23,* 296–326. doi:10.1080/01690960701571570

Creel, S. C., & Dahan, D. (2010). The effect of the temporal structure of spoken words on paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 110–122. doi:10.1037/a0017527

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

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,* 559–597. doi:10.1017/S0305000906007471

Dryer, M. S. (2005). Prefixing versus suffixing in inflectional morphology. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 110–116). New York, NY: Oxford University Press. Retrieved from http://wals.info/feature/26

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (Vol. 3, pp. 1393–1396). New York, NY: Institute of Electrical and Electronics Engineers.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York, NY: Teachers College, Columbia University. (Original work published in 1885) Retrieved from http://psychclassics.yorku.ca/Ebbinghaus

Eddington, D. (2009). Spanish verbal inflection: A single- or dual-route system? *Linguistics, 47,* 173–199. doi:10.1515/LING.2009.006

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105,* 247–299. doi:10.1016/j.cognition.2006.09.010

Endress, A. D., Cahill, D., Block, S., Watumull, J., & Hauser, M. D. (2009). Evidence of an evolutionary precursor to human language affixation in a non-human primate. *Biology Letters, 5,* 749–751. doi:10.1098/rsbl.2009.0445

Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition, 13,* 483–495.

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology, 62,* 2187–2209. doi:10.1080/17470210902783646

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 235–250. doi:10.1037/a0017164

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13,* 348–353. doi:10.1016/j.tics.2009.05.005

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134,* 406–419. doi:10.1037/0096-3445.134.3.406

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140. doi:10.1037/0096-3445.127.2.107

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635. doi:10.1016/S0022-5371(73)80042-8

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436. doi:10.1111/1467-9280.00476

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7,* 567–580. doi:10.1111/j.1467-7687.2004.00381.x

Garvey, C. J. (1964). *A sketch of Malagasy grammar.* Washington, DC: Center for Applied Linguistics.

Gerken, L., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development, 4,* 228–248. doi:10.1080/15475440802143117

Greenberg, J. (1957). *Essays in linguistics.* Chicago, IL: University of Chicago Press.

Hahn, U., & Nakisa, R. C. (2000). German inflection: Single route or dual route? *Cognitive Psychology, 41,* 313–360. doi:10.1006/cogp.2000.0737

Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition, 56,* 61–98. doi:10.1016/0010-0277(94)00655-5

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes, 10,* 601–630. doi:10.1080/01690969508407115

Henson, R. N. A. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology, 36,* 73–137. doi:10.1006/cogp.1998.0685

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27,* 915–927.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1,* 151–195. 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,* 30–66. doi:10.1016/j.cogpsych.2009.01.001

Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes, 21,* 394–431. doi:10.1080/01690960400013213

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, USA, 96,* 7592–7597. doi:10.1073/pnas.96.13.7592

Julien, M. (2002). *Syntactic heads and word formation.* New York, NY: Oxford University Press.

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., & Martens, E. (2007). Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology, 54,* 283–318. doi:10.1016/j.cogpsych.2006.07.002

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. 1, pp. 131–175). Dordrecht, the Netherlands: Foris.

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

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23,* 1089–1132. doi:10.1080/01690960802193688

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 876–895. doi:10.1037/a0013484

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* Cambridge, MA: MIT Press.

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

Marslen-Wilson, W. D., & Tyler, L. K. (1997). Dissociating types of mental computation. *Nature, 387,* 592–594. doi:10.1038/42456

Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2,* 428–435. doi:10.1016/S1364-6613(98)01239-X

Miceli, G., Capasso, R., & Caramazza, A. (1994). The interaction of lexical and sublexical processes in reading, writing and repetition. *Neuropsychologia, 32,* 317–333. doi:10.1016/0028-3932(94)90134-1

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition, 94,* 215–240. doi:10.1016/j.cognition.2004.02.002

Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science, 10,* 237–254. doi:10.1111/j.1467-7687.2007.00582.x

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. doi:10.1037/0033-295X.101.1.53

Orlov, T., Yakovlev, V., Hochstein, S., & Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. *Nature, 404,* 77–80. doi:10.1038/35003571

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 548–568. doi:10.1037/0278-7393.21.3.548

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43,* 93–120. doi:10.1111/j.1467-1770.1993.tb00174.x

Pinker, S. (1991). Rules of language. *Science, 253,* 530–535. doi:10.1126/science.1857983

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York, NY: Basic Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–193. doi:10.1016/0010-0277(88)90032-7

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition, 38,* 43–102. doi:10.1016/0010-0277(91)90022-V

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21–69. doi:10.1016/0010-0277(93)90057-3

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

Quinn, P. C., & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science, 16,* 511–515. doi:10.1111/j.0956-7976.2005.01567.x

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology, 45,* 45–94. doi:10.1016/S0010-0285(02)00001-4

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37,* 118–139. doi:10.1006/jmla.1997.2510

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599–1603. doi:10.1126/science.275.5306.1599

Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43,* 195–201. doi:10.1037/h0054636

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263–272.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638–647. doi:10.1016/S0022-5371(75)80051-X

Terrace, H. S. (2005). The simultaneous chain: A new approach to serial learning. *Trends in Cognitive Sciences, 9,* 202–210. doi:10.1016/j.tics.2005.02.003

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4,* 156–163. doi:10.1016/S1364-6613(00)01462-5

Tyler, L. K., Marslen-Wilson, W. D., & Stamatakis, E. A. (2005). Differentiating lexical form, meaning, and structure in the neural language system. *Proceedings of the National Academy of Sciences, USA, 102,* 8375–8380. doi:10.1073/pnas.0408213102

Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63,* 1012–1020. doi:10.2307/1131250

Westermann, G., & Plunkett, K. (2007). Connectionist models of inflection processing. *Lingue e linguaggio, 6,* 291–311.

Received September 8, 2009
Revision received April 26, 2010
Accepted April 27, 2010 ■