The Relationship Between Frequency and Irregularity in the Evolution of Linguistic Structure: An Experimental Study

Kenny Smith (kenny.smith@ed.ac.uk) & Clem Ashton
Centre for Language Evolution, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD

Helen Sims-Williams
Surrey Morphology Group, School of Literature and Languages, University of Surrey, Guildford, GU2 7XH

Abstract

The expressive power of natural languages depends on their regular compositional structure, which allows us to express and understand an infinite set of messages. However, a complete model of language evolution should also account for irregular exceptions to regular rules, common in natural languages. Historical linguistics has established a correlation between irregularity and frequency in language use, which has been attributed to preferential irregularisation of frequent items, or preferential regularisation of infrequent items. In an iterated learning experiment where participants learn and reproduce a miniature language across multiple generations, we show that this correlation can be explained by the relationship between frequency, regularity and learnability, without needing to appeal to frequency-dependent irregularisation. We find that systems of plural marking regularise across generations of transmission, but that high-frequency items remain irregular. Our results further show that the persistence of irregularity is due to high frequency overriding pressures which normally reduce learnability, such as low generalisability of the inflectional strategy (suppletion is disfavoured except in high frequency items) and low type frequency (belonging to a small inflectional class is disfavoured except in high frequency items).

Keywords: language evolution; iterated learning; artificial language learning; frequency; irregularity

Introduction

Compositional structure is a defining feature of human language: we build complex expressions using a finite set of morphemes/words and a grammar which dictates how those building blocks can be combined, with the form of a complex expression being a predictable consequence of the component parts. Compositionality therefore enables the expression of novel concepts that characterises the open-ended nature of human communication (e.g. Hockett, 1960). Compositionality can emerge purely as a consequence of biases inherent in the cycle of learning and use by which natural languages persist, as has been demonstrated both in simulation (e.g. Kirby, 2002; Smith, Tamariz, & Kirby, 2013) and in subsequent experiments where human participants learn, use and transmit artificial languages (e.g. Kirby, Cornish, & Smith, 2008; Beckner, Pierrehumbert, & Hay, 2017; Kirby, Tamariz, Cornish, & Smith, 2015).

Less attention has been paid to evolutionary explanations for the presence of irregularity in natural language. Although there is cross-linguistic variation in its prevalence (Kiefer, 2000), all known languages feature some amount of irregularity (Stolz, Otsuka, Urdze, & van der Auwera, 2012). Irregularity is present both in morphology and syntax, and exists on a scale (Kroch, 1989) from highly regular and productive (e.g. the regular past tense suffixation seen in *jump–jumped*), to semi-productive (e.g. *sing–sang*, *begin–began*), to suppletive marking, which can only apply to a single item (e.g. *go–went*). If human learning biases favour regular rules, why do these irregularities persist so stubbornly?

There is a well-known correlation between frequency and irregularity (e.g. Bybee, 1991; Leech, Rayson, & Wilson, 2002; Wu, Cotterell, & O'Donnell, 2019): irregulars tend to be highly frequent in their usage, and low frequency expressions tend to follow regular patterns. In historical linguistics, the explanation for the correlation between frequency and irregularity appeals to the interplay of processes of language change that introduce regularity and irregularity (Pagel, Atkinson, & Meade, 2007; Bybee, 1995; Sims-Williams, 2022). Two of these interacting processes are sound change and analogy. Sound change tends to create irregularity by introducing meaningless alternations of form, as when vowel shortening in closed syllables affected the paradigm of Old English *me:tan* 'to meet', *me:tte* 'met', and other reductive sound changes removed the conditioning environment that made this phonological variation predictable, giving us the irregular vowel alternation seen in English today *meet* [iː], *met* [ɛ]. On the other hand, analogical changes often restore the regularity disrupted by sound change, by extending marking strategies found elsewhere in the language (Campbell, 2013): compare the inherited form *dreamt*, which has been affected by the same pre-consonantal shortening as *kept*, with its analogical variant *dreamed*, produced using the regular rule for past suffixation.

The correlation between frequency and irregularity may be because frequent items are more susceptible to irregularisation, or because they are more resistant to regularisation, or both. There is evidence that frequent expressions are more susceptible to sound change, especially the kinds of reductive sound changes that introduce irregularity (Garrett, 2015; Bybee, 2017; Todd, Pierrehumbert, & Hay, 2019). If so, we might expect more irregularities to accumulate in the paradigms of highly frequent lexemes, in the same way as for Old English *me:tan*. There is also evidence that frequent items are more resistant to regularisation via analogy (Bybee, 1995; Sims-Williams, 2022). As such, high frequency can be viewed as both a driving force for the introduction of irregularity, and a protective force against regularisation.

Computational models of language learning and transmission have shown that the correlation between frequency and irregularity can in principle arise purely from protection from analogical regularisation associated with high frequency, without appealing to preferential irregularisation of high-frequency items (Kirby, 2001; Kirby, Dowman, & Griffiths, 2007; Morgan & Levy, 2016; Cuskley et al., 2017; Liu & Morgan, 2020). For instance, in a Bayesian model where learner priors favour regularity (Kirby et al., 2007), the prior is outweighed by the data-dependent likelihood for high frequency items, allowing high frequency irregulars to escape regularisation pressure. Here we use an iterated artificial language learning paradigm to test experimentally whether similar results are found with human language learners. Previous experimental work in iterated learning has looked at the conditions under which regular systems evolve (e.g. Kirby et al., 2008, 2015; Smith & Wonnacott, 2010), and other work focussing on language learning in individuals has explored how non-uniform frequencies can impact on word segmentation (Lavi-Rotbain & Arnon, 2022) or word learning under referential uncertainty (Hendrickson & Perfors, 2019). We test whether a language in which different items appear at different frequencies during learning will, over repeated generations of learning and use, develop the frequency-irregularity correlation seen in natural languages. Our data also allows us to determine whether, in our experiment, this correlation arises from preferential irregularisation of frequent items, preferential regularisation of infrequent items, or both.

Method

The experiment follows the iterated artificial language learning paradigm from Kirby et al. (2015) (see Figure 1): participants were organised into transmission chains of multiple generations, where each generation consists of pair of participants who are trained on the same language and then play a communication game, taking turns to produce descriptions for their partner, and attempting to identify images based on the description provided by their partner. The language produced during interaction by one of the participants at generation n is then used as the target language in training the generation $n+1$ pair. Iterated learning (the passing of the language from generation to generation) allows the languages to be shaped by the cumulative effects of learning and use; enforcing a communicative pressure prevents the languages from collapsing to a degenerate state where no distinctions are encoded at all. We manipulated the frequency with which objects and their descriptions appear during training: in the Uniform condition all objects were shown the same number of times during training, in the Skewed condition some objects were shown more than others.

Participants

We ran 40 chains (20 chains per condition) for 5 generations per chain (400 participants). Participants were recruited via Prolific, and were native speakers of English who had a high approval rating and had already completed 10+ studies on Prolific. The experiment took approximately 35 minutes. Participants were paid £5 plus a bonus of up to £0.96.

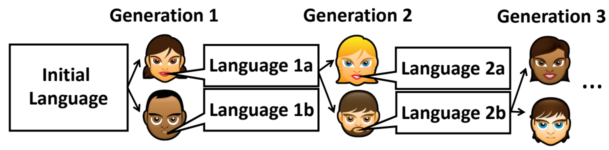

Figure 1: The iterated learning procedure.

Stimuli

Participants were tasked with learning an artificial language which provided labels for 6 unfamiliar objects, each appearing in singular and plural form (see Figure 2). Number was chosen as salient dimension of variation, which participants would be willing to mark using inflection (as they doin other artificial language experiments, e.g. Smith & Wonnacott, 2010). The objects were selected from the NOUN Database (Horst & Hout, 2016).

Initial languages

40 initial languages (one per chain) were generated to teach to participants at generation 1. The initial languages had an irregular structure where each object used a unique strategy to mark plurality (see Figure 2). Labels for singular scenes were two-syllable stems. Three objects had suppletive plural forms: a two-syllable form, unrelated to the singular. Three objects used suffixation for plurals: the plural was composed of the singular stem and a unique, 1-syllable suffix (i.e. a different suffix for each object). This configuration was designed to facilitate the emergence of fully regular inflection (e.g. through one of the suffixes spreading to mark plural for all objects), whilst also making resistance to regularisation possible (e.g. if suppletive forms remained in the language, or if each object persisted with an idiosyncratic suffix).

The labels were built from an inventory of 12 syllables: *nu, wo, za, sla, mo, vi, bli, hu, shru, ri, dra, plo*. A set of labels was randomly generated in accordance with the structure described above, with several additional constraints: each label in the language had to start with a unique syllable; there could be no repeated syllables within a label; syllables used as suffixes were not used anywhere else in the labels. These constraints were imposed to increase label distinctiveness and to avoid participants inferring structure beyond the 3 suffixes built into the initial language.

Procedure

The experiment was coded in JavaScript. Real-time interaction between crowdsourced participants was achieved using WebSocket connections to a Python server which paired and coordinated participants, and iteration was automated using an SQL database which tracked active chains and allocated new pairs to open chains.

	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6
Singular image						
Singular label	viza	drashru	wodra	mowo	shrunu	plonu
Plural image						
Plural label	zawo	drashru**bli**	huvi	mowo**sla**	nuplo	plonu**ri**
Number of trials in uniform training	16	16	16	16	16	16
Number of trials in skewed training	24	24	12	12	12	12

Figure 2: Example of an initial language, plus frequencies used in training. Singular scenes have random, two-syllable labels. Plural scenes have either two-syllable, suppletive labels (in red) or a unique suffix on the singular (in bold).

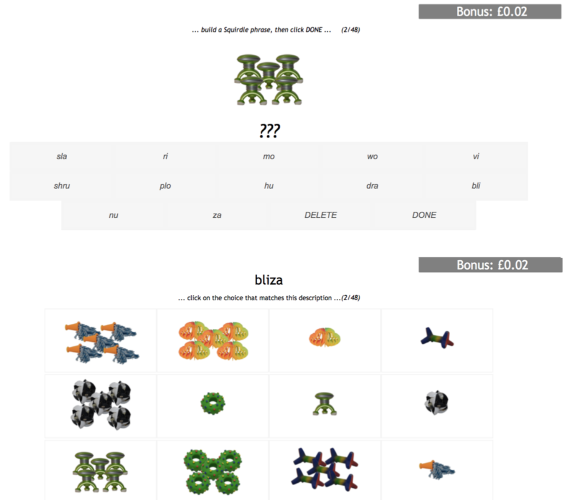

Figure 3: Example of the communication game: Upper: the sender is shown a scene and must build a label using the 12 syllable buttons. Lower: the receiver receives the label from the sender and must choose which image it refers to.

Training After completing a short introductory series of trials intended to deter or eliminate unengaged participants before pairing, participants entered a waiting room and were paired with another participant. Each pair was assigned a language to learn, and started training. There were 4 training blocks (24 trials each), and 2 short interim test blocks (4 trials each). The first training block consisted of passive exposure trials (an image presented with its label for 4 seconds), designed to familiarise participants with the language. The following 3 blocks consisted of label selection trials (participants were shown an image and had to select the correct label from a choice of 4 labels; after selection they received feedback, and the correct label appeared for 4 seconds). After label selection blocks 1 and 2 participants were presented with 4 interim test trials, where they were shown a randomly-selected image and asked to build its label by clicking on syllable buttons, choosing from all 12 syllables in the inventory. Participants were not given any feedback during interim test trials. The interim test trials were intended to familiarise participants with the process of producing labels. Syllable buttons were used rather than free typing in order to prevent participants from straying towards producing English or near-English labels.

Communication After both participants finished training, they played a communication game (see Figure 3). In each communication trial the sender was shown a image and used the syllable buttons to construct a label to send to the receiver. The receiver saw the label and had to select which image it referred to, choosing from all 12 possible images, encouraging senders to encode both object and number in their descriptions. After each trial, both participants were shown feedback on whether they were correct / incorrect, alongside the sender's label, the target image, and the image chosen by the receiver. Participants received a £0.02 bonus for each successful trial to encourage effort and engagement. Participants alternated the roles of sender and receiver, each producing a label for every scene in the language (in a randomised order) twice, yielding a total of 48 communication trials.

Frequency manipulation in training (see Figure 2) In the Uniform condition, participants encountered each object 16 times over the 96 trials in training: 8 as a singular, 8 as a plural. We refer to this as Mid frequency in the analyses below. In the Skewed condition, the 6 objects were split into High frequency (2 objects) and Low frequency (4 objects). High frequency objects were shown 24 times each (12 singular, 12 plural), and low frequency objects were shown 12 times each (6 singular, 6 plural). In the initial language, one high-frequency and two low-frequency objects had suppletive forms; one high-frequency and two low-frequency objects had suffixing plural forms. Note that the frequency skew did not apply in the interim test trials or in the communication phase.

Iteration When a pair finished the communication game, an output language was formed using the labels produced by one participant as sender on their second pass through the set of 12 scenes. This output language was used as the training language taught to the pair at the next generation of that chain. A minimum accuracy threshold had to be met in order for languages to be used for iteration: the labels for the 6 singular scenes had to have an average learning accuracy above 0.6, where learning accuracy is inverse normalised edit distance between the label from the training language and the label the participant produced, i.e. $1-[L(i,j)/max(|i|,|j|)]$ where $L(i,j)$ is the Levenshtein distance between trained label i and produced label j and $|i|$ is the length of label i; learning accuracy 1 indicates perfect reproduction of the training label, 0 indicates no correspondence. This filter was applied to exclude participants who were not able to reproduce their input language with a reasonable degree of accuracy. The accuracy threshold only applied to singular labels so as to allow innovation in the plural marking system. If both participants in a pair met the threshold we selected a participant to iterate from at random; if one member of a pair met the accuracy threshold we iterated from their data; if neither member of a pair met the threshold then we discarded that pair and re-ran that generation. 11 pairs were rejected from iteration due to both participants falling below the accuracy threshold. There were 35 pairs where only one participant fell below the accuracy threshold.

Measuring (ir)regularity via inflectional class size

Class size measures how many objects in the language use a particular strategy to mark plurality. Suppletive plurals always belong to a class of size 1, because suppletive marking by definition can only apply to a single item. A plural formed by suffixation can belong to a larger class if the same suffix is used to form other plurals. In the initial languages there were 6 classes, each of size 1 (3 suppletive classes and 3 suffixing classes, each using a different suffix). In this experiment a perfectly regular system would consist of a singular class (e.g. suffix -nu) with 6 members (i.e. all 6 objects). Our analysis focuses on the size of the class each item belongs to[1], the type of inflection it uses (suppletive or suffixing[2]), and learning accuracy (as defined above).

Results

We predicted that regularity would increase over generations, in line with other iterated learning experiments showing the emergence of regularity (e.g. Smith & Wonnacott, 2010; Kirby et al., 2015; Beckner et al., 2017). We also predicted that irregularity would be associated with highly frequent items. More specifically, we predicted that (1) a regular system of inflection would emerge in both the Uniform and Skewed conditions, but in the Skewed condition the high frequency suppletives will survive for longer; (2) high frequency suffixing items would be the most likely to absorb new items into their class, therefore forming the basis for the regular class.

Figure 4 (left) plots class size against generation and frequency (4 levels: Mid frequency as seen in the Uniform chains, Low and High frequency as in the Skewed chains; within High frequency items we distinguish based on their initial plural marking strategy since this is informative about the origins of the regular class). Class size clearly increases over generations as larger regular inflectional classes form; however, class size remains lower in high-frequency suppletive items, matching our prediction that irregularity would persist for those items. The fact that class size increases as much for the high-frequency suffixing items as for lower frequency items suggests either that (1) the emerging regular class forms from that initial high-frequency suffix, or (2) these high frequency items (but not the high frequency suppletives) join a regular class which has its origins elsewhere, and high frequency therefore only provides protection from regularisation for suppletive items. Our learnability analysis below speaks against the second possibility, in that high frequency suffixing and suppletive items are learnt with similar accuracy. An analysis of the origins of the regular inflections in the 18 Skewed chains which developed a clear majority regular inflection shows that the eventual regular is usually (12 times of 18) seen first on a high-frequency item, supporting the first explanation.

We used a linear mixed effects model to analyse class size with generation, frequency and their interaction as fixed effects.[3] We used Helmert coding for frequency, such that the three levels of the frequency fixed effect indicate 1) Mid vs Low, 2) High Suffixing vs the mean of Mid and Low, and 3) High Suppletive vs the mean of High Suffixing, Mid and Low. The model shows a significant effect of generation (b=0.24, SE=0.04, $p < .001$), indicating that class size increases over generations, and an interaction between generation and the final contrast for frequency (b=-0.04, SE=0.01, p= .003), indicating that the increase in class size over generations for High Suppletive forms is lower. All other effects and interactions are n.s. ($p>0.27$), including the frequency by generation interaction that would indicate that the High Suffixing forms remain more irregular than Low/Mid items.[4]

In order to explore the mechanisms driving this pattern of regularity and irregularity, we conduct several analyses of factors influencing label length and learning accuracy (as defined above, i.e. inverse normalised edit distance).

First we check whether high-frequency items are preferentially shortened (or kept short) during interaction: based on other artificial language paradigms which feature multiclick production procedures like ours (Fedzechkina & Jaeger, 2020), we might expect that participants preferentially shorten high-frequency items to minimise production effort

[1]We therefore operationalise the regularity-irregularity continuum using type frequency: items with higher type frequency, i.e. belonging to a larger inflectional class, are more regular. See Herce (2019) for discussion of other conceptions of (ir)regularity.

[2] In addition to suffixing and suppletion, we saw a small number of other strategies for plural formation: prefixing, prefixing plus suffixing, stem alternation (part of the stem altered to form plural), or non-alternation (plural same as singular). Since these strategies, like suffixing but unlike suppletion, can be generalised across multiple objects we include them with suffixing in the plots and learnability analyses that follow; analyses which exclude these inflectional strategies and focus solely on suppletion versus suffixing produce the same patterns of results.

[3]Models were run in R (R Core Team, 2019) using lmer (Bates, Mächler, Bolker, & Walker, 2015); plots were produced in ggplot2 (Wickham, 2009). The random effects structure for the class size analysis consists of by-chain random slopes for generation; models with by-chain random slopes for frequency did not converge but produced a similar pattern of significant effects.

[4]One possibility that this analysis does not rule out is that initially-suppletive items increase in class size more slowly that initially-suffixing forms, regardless of frequency. An analysis with generation, frequency (Helmert-coded to produce two contrasts: Low vs Mid; High vs the mean of Mid and Low), and initial inflection type (suffix; suppletive) shows a significant interaction between generation and initial inflection type (b=-0.06, SE=0.01, $p<.001$), indicating that suppletives do regularise more slowly, but also a three-way interaction between generation, initial inflection type and frequency (b=-0.02, SE=0.01, p=.044), indicating that high frequency suppletives regularise more slowly still.

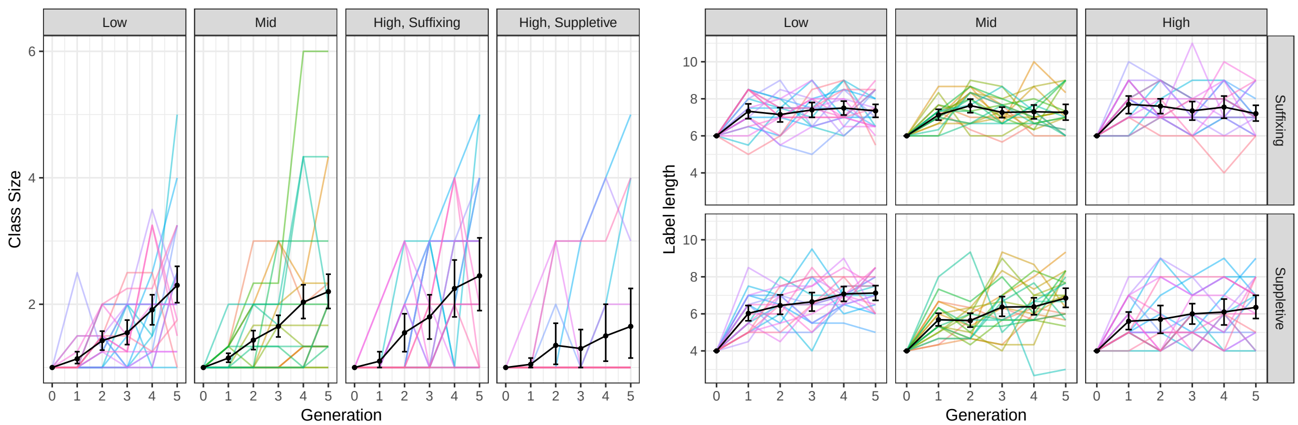

Figure 4: Left: Inflectional class size versus generation (initial language shown as generation 0), for Low, Mid and High frequency nouns, with High frequency items split by inflectional type in the initial language. Right: The effect of generation, frequency and initial inflectional strategy on label length (in characters). Black lines give means and 95% confidence intervals, coloured lines show average class size or length in individual chains.

(Kanwal, Smith, Culbertson, & Kirby, 2017), which would favour irregular suppletive forms for high-frequency items (although recall that frequency was uniform in interaction). Figure 4 (right) plots label length against generation. We analysed this data using a linear mixed effects model with generation, initial inflection type (suffixing or suppletive, sum-coded), frequency (Helmert-coded giving two contrasts: Low vs Mid; High vs the mean of Mid and Low) and their interaction as fixed effects.[5] The model shows a significant effect of generation (b=0.33, SE=0.03, p<.001), indicating that labels generally increase in length, and a significant interaction between initial inflection type and generation (b=0.14, SE=0.02, p<.001), indicating that initially-suppletive plurals increase in length more rapidly; however, all interactions which would indicate that high-frequency items or high-frequency suppletive items are an exception to this tendency for increasing length are not significant (smallest p=.139), indicating that the frequency-irregularity correlation we see is unlikely to be due to preferential shortening of high-frequency forms.

Next we turn to learning-based mechanisms. Figure 5 (left) plots the accuracy with which suffixing versus suppletive forms are learned. We analyse this learning accuracy data with a linear mixed effects model with inflection type (sum-coded, suffixing vs suppletive; see footnote 2), frequency (Helmert-coded, 2 contrasts) and their interaction as fixed effects.[6] The model shows a significant effect of inflection type (b=-0.06, SE=0.008, p< .001), indicating worse learning of suppletives, and positive effects for both frequency factors (Low to Mid: b=0.03, SE=0.01, p=.008; Low/Mid to High: b=0.03, SE=0.007, p<.001) indicating better learning of higher frequency forms. There is a marginal interaction between the first frequency contrast and inflection type (b=0.02, SE=0.009, p=.067) and a significant interaction between the second frequency factor and inflection type (b=0.02, SE=0.006, p=0.019), indicating that Mid (marginally) and High (more clearly) frequency reduces the learnability penalty associated with suppletive inflection.

Figure 5 (right) plots the accuracy with which plural forms are learned, given the size of inflectional class they belong to. We again run a linear mixed effects model analysing learning accuracy with class size (centred class size such that the model intercept indicates class size 1), frequency (Helmert coded, 2 contrasts) and their interaction as fixed effects. The model shows a significant effect of both frequency contrasts (Low to Mid: b=0.04, SE=0.01, p=.002; Low/Mid to High: b=0.04, SE=0.008, p<.001), indicating that higher frequency improves learnability of the idiosyncratic class size 1 plurals. There is a positive effect of class size (b=0.07, SE=0.02, p< .001), indicating that items belonging to a larger class are learned more accurately, but interactions between frequency and class size indicate this advantage of class size reduces for more frequent items (significant interaction between the first frequency contrast and class size, b=-0.02, SE=0.01, p=.047; the interaction between the second frequency contrast and class size is negative but n.s., b=-0.01, SE=0.01, p=.182). As with the analysis of suppletive inflection, the effect of frequency for class size 1 items shows that high-frequency items escape the learnability penalty normally associated with belonging to a small inflectional class.[7]

[5]The random effects structure for this model consisted of a by-chain random slope for generation; models with random slopes for frequency and/or initial inflection type produced convergence warnings but showed the same pattern of significant effects.

[6]The random effects structure for this model and the equivalent model with class size consisted of a by-participant random intercept; models with by-participant random slopes for frequency and/or class type did not converge, presumably due to sparsity, but showed the same pattern of significant effects.

[7]We cannot determine whether inflectional type and class size

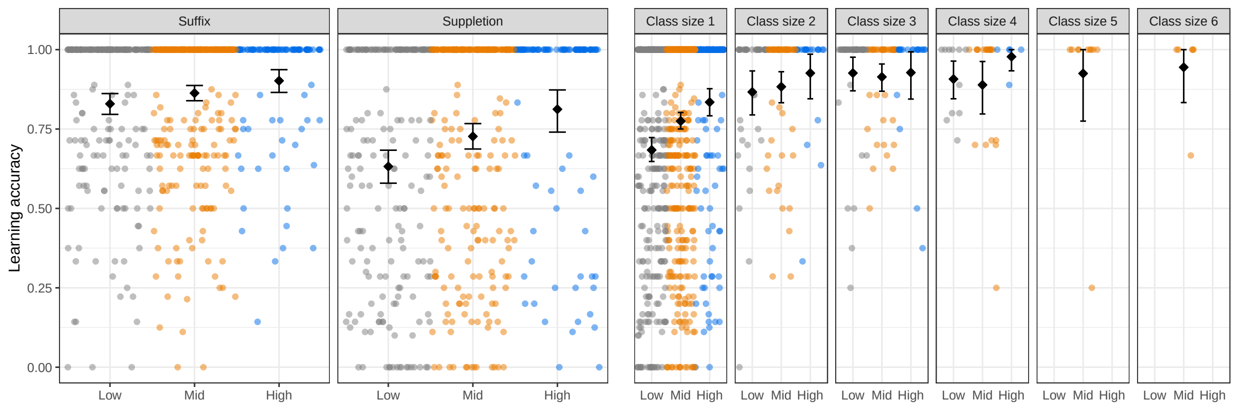

Figure 5: The effect of frequency and inflectional type (left) and inflectional class size (right) on learning accuracy. Each point represents the accuracy of a production by a participant during the communication phase of the experiment (data from all generations plotted here), with points jittered to improve readability; black diamonds and error bars give means and 95% CIs.

Discussion

Our results showed an overall tendency for class size to increase over generations as plural marking regularises. High frequency suppletives resisted regularisation for longer than other forms, supporting our hypothesis that high frequency items are more resistant to regularisation. The high frequency suffixing items, on the other hand, increased in class size at a similar rate to other items. This is also consistent with our hypothesis, since we expect greater resistance to regularisation in these items to be counteracted by a higher probability that they will attract other items into their inflectional class.

Pressures from learning can explain both the overall tendency towards regularisation, and why this tendency is more pronounced in infrequent items. Participants are less likely to remember which plural marking strategy low-frequency items employ. When called upon to produce these plurals, they are therefore more likely to extend an alternative marking strategy, which is more likely to come from a larger class, since marking strategies employed by more items are better evidenced in the training data. This interpretation is supported by our analysis of learning accuracy, which showed a learning penalty for both suppletives and low type frequency (small class size) items, in both cases a penalty which is attenuated for items with high frequency.

The correlation between frequency and irregularity in natural languages could be due to preferential irregularisation of frequent forms, preferential regularisation of infrequent forms, or both. The mechanism that has been proposed for preferential irregularisation of frequent forms is reductive sound changes, which may be more likely to target highly frequent or predictable items. In our data the general trend is for both class size and label length to increase over generations, with no significant difference in lengthening rate across the frequency bands. Therefore we found no evidence for preferential reduction of frequent forms leading to irregularity, but nonetheless saw the emergence of the correlation between frequency and irregularity that is also found in natural languages. While this by no means rules out a role for preferential reduction of frequent forms in language change, it does show that it is not necessary for a frequency-irregularity correlation to emerge.

contribute independently to learnability, since class size is by definition 1 for suppletives. However, repeating the analysis of the effect of class size on learnability while excluding suppletives still shows a positive effect of class size (b=0.04, SE=0.01, $p<.001$); an analysis of all class size 1 items (fixed effects of frequency and inflection type) shows that suppletive forms are generally learned slightly less accurately than suffixing forms (b=-0.05, SE=0.01, $p<.001$); taken together, these results suggest inflectional type and class size both contribute to learnability.

Acknowledgments

This research received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 681942, held by K. Smith).

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1).

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, *2*, 160–176.

Bybee, J. (1991). Natural morphology: The organization of paradigms and language acquisition. In T. Huebner & C. A. Ferguson (Eds.), *Language Acquisition and Language Disorders* (Vol. 2, p. 67). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/lald.2.08byb

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

Bybee, J. (2017). Grammatical and lexical factors in sound change: A usage-based approach. *Language Variation and Change*, *29*(3), 273–300. doi: 10.1017/S0954394517000199

Campbell, L. (2013). *Historical linguistics: An introduction* (3rd ed.). Edinburgh: Edinburgh University Press.

Cuskley, C., Castellano, C., Colaiori, F., Loreto, V., Pugliese, M., & Tria, F. (2017). The regularity game: Investigating linguistic rule dynamics in a population of interacting agents. *Cognition*, *159*, 25–32. doi: 10.1016/j.cognition.2016.11.001

Fedzechkina, M., & Jaeger, T. F. (2020). Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition*, *196*, 104115. doi: 10.1016/j.cognition.2019.104115

Garrett, A. (2015). Sound change. In *The Routledge handbook of historical linguistics.* Routledge.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition*, *189*, 11–22. doi: 10.1016/j.cognition.2019.03.005

Herce, B. (2019). Deconstructing (ir)regularity. *Studies in Language*, *43*, 44–91. doi: 10.1075/sl.17042.her

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication* (pp. 392–430). Washington, DC: American Institute of Biological Sciences.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. doi: 10.3758/s13428-015-0647-3

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52. doi: 10.1016/j.cognition.2017.05.001

Kiefer, F. (2000). Regularity. In G. Booij, C. Lehmann, J. Mugdan, W. Kesselheim, & S. Skopeteas (Eds.), *Morphologie* (pp. 296–302). Walter de Gruyter. doi: 10.1515/9783110111286.1.4.296

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110. doi: 10.1109/4235.918430

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models* (pp. 173–203).

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686. doi: 10.1073/pnas.0707835105

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, USA*, *104*, 5241–5245.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and Communication in the Cultural Evolution of Linguistic Structure. *Cognition*, *141*, 87–102.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244. doi: 10.1017/S0954394500000168

Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, *223*, 105038. doi: 10.1016/j.cognition.2022.105038

Leech, G. N., Rayson, P., & Wilson, A. (2002). *Word frequencies in written and spoken English: Based on the British National Corpus* (Nachdr. ed.). Harlow Munich: Longman.

Liu, Z., & Morgan, E. (2020). Frequency-dependent Regularization in Constituent Ordering Preferences. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

Morgan, E., & Levy, R. (2016). Frequency-dependent regularization in iterated learning. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).*

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*, 717–720.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sims-Williams, H. (2022). Token frequency as a determinant of morphological change. *Journal of Linguistics*, *58*(3), 571–607. doi: 10.1017/S0022226721000438

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1348–1353). Toronto: Cognitive Science Society.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. doi: 10.1016/j.cognition.2010.06.004

Stolz, T., Otsuka, H., Urdze, A., & van der Auwera, J. (2012). Introduction: Irregularity – glimpses of a ubiquitous phenomenon. In T. Stolz, H. Otsuka, A. Urdze, & J. van der Auwera (Eds.), *Irregularity in Morphology (and beyond)* (pp. 7–37). Akademie Verlag. doi: 10.1524/9783050059587.7

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20. doi: 10.1016/j.cognition.2019.01.004

Wickham, H. (2009). *Ggplot2*. New York, NY: Springer New York.

Wu, S., Cotterell, R., & O'Donnell, T. J. (2019). Morphological Irregularity Correlates with Frequency. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5117–5126).