UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Testing the Effects of the Implicative Structure and Noun Class Size on the Learnability of Inflectional Paradigms in Adults and Artificial Neural Networks

Permalink

https://escholarship.org/uc/item/1hj6k6r5

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Johnson, Tamar
Elsner, Micha
Smith, Kenny

Publication Date

2024

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Testing the Effects of the Implicative Structure and Noun Class Size on the Learnability of Inflectional Paradigms in Adults and Artificial Neural Networks

Tamar Johnson (t.johnson@uva.nl)
Institute for Logic, Language and Computation, University of Amsterdam

Micha Elsner (elsner.14@osu.edu)
Department of Linguistics, The Ohio State University

Kenny Smith (Kenny.Smith@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh

Abstract

Variation in inflectional morphology across languages raises questions about the factors affecting their learnability. This study explores the effects of two suggested factors: the implicative structure of the paradigm and the distribution of forms within it, and how they interact to affect the learnability of the system. Our results from a human behavioral study and artificial neural network simulations suggest that these factors influence learning, though type frequency may only serve as a proxy for the effects of token frequency.

Keywords: implicative structure; inflectional morphology; learnability; artificial neural networks

Introduction

There is a wide variation across languages in how nouns are marked for grammatical information, i.e., their inflectional morphology. In Mandarin, for example, nouns are not marked at all for grammatical information, whereas Arabic uses inflectional morphology to mark dozens of grammatical functions (including e.g. number, gender, case, person, mood, tense, and voice). Several measures and methods are proposed in the literature to quantify the complexity of inflectional morphology. Some measure the number of forms which mark the same grammatical information, others the distribution across forms, while another approach models the implicative structure of paradigms. According to the latter approach, inflectional systems where forms can be predicted by analogy to other known forms are easier to learn and process. Typological studies show that the complexity of inflectional systems in natural languages occupy quite a limited range with respect to implicative structure, suggesting it as an organizing principle in diachronic change (e.g., Ackerman & Malouf, 2013; Sims-Williams, 2016). Recent behavioural studies have tested the effect of implicative structure on inflectional paradigm learning and found its effect to be secondary to the sheer number of forms in the paradigm (Johnson, Gao, Smith, Rabagliati, & Culbertson, 2021; Johnson, Culbertson, Rabagliati, & Smith, 2020). Here we test how another factor, the frequency distribution over forms in the paradigm (specifically, whether classes have many or few members), affects paradigm learning, and how frequency interacts with the implicative paradigmatic structure

Ackerman and Malouf (2013, 2015) describe the implicative complexity (I-complexity) of an inflectional system as the average uncertainty which a speaker has about one form of a word given another— for example, uncertainty about the past participle of a verb (*goed* or *gone*) given the past form (*went*). High I-complexity has been argued (Cotterell, Kirov, Hulden, & Eisner, 2019) to represent a barrier to learning, but artificial language learning experiments (Johnson et al., 2021, 2020) have shown a relatively modest effect of I-complexity on learning rate for human participants. The artificial languages studied in Johnson et al. (2021, 2020) and earlier work (Seyfarth, Ackerman, & Malouf, 2014), however, use lexicons in which words and inflection classes have balanced frequencies (i.e. contain the same number of lexical items, each occurring with the same frequency in learning). This means that learners must master the implicative relationships for all classes in order to inflect correctly.

As Sims and Parker (2016) point out, I-complexity on its own does not fully characterize the challenge of learning an inflection system, because it neglects the role of type frequency. Many languages exhibit some degree of inflectional irregularity; they have inflection classes with few members (low type frequency), though those members are themselves often highly frequent (i.e. have high token frequency; Bybee, 1995). In systems with irregular classes, certain words can be learned as lexical exceptions (like English *be*). Moreover, irregulars often contribute in an outsized way to the I-complexity of the system they inhabit, a tendency which Stump and Finkel (2013) call "marginal detraction."

Here we report an artificial language learning experiment with human participants which shows effects of the implicative structure of the paradigm and token frequency, qualifying the original claim that I-complexity directly measures the difficulty of learning a morphological system. These results also support the claims made in Sims and Parker (2016); Parker and Sims (2020) that the marginal detraction property supports the learning of inflectional paradigms, though in a way that matches token frequency effects on learnability. We simulate the effects of learning a larger lexicon using neural networks as surrogates for humans; scaled-up lexicons do not show marginal detraction effects, potentially setting limits on how frequent an inflection class must be to qualify as 'marginal'.

Behavioural Experiment

Following the method from Johnson et al. (2020), we had crowdsourced participants attempt to learn inflectional

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

paradigms in an artificial language, where suffixes of nouns indicated number. We manipulate the I-complexity of the target paradigms and the frequency distribution over inflectional classes.

Methods

Target Paradigms The artificial language consist of nouns inflected for three grammatical numbers according to one of two inflectional paradigms, manipulating the implicative structure of the paradigm (as in Johnson et al., 2020). The language's lexicon consist of eighteen CVCV nouns. Stems in the artificial language are randomly paired with meanings (images of objects and animals) and assigned to one of three noun classes.

Inflectional markers are seven CVC monosyllabic suffixes (*-fel, -fob, -fir, -fam, -fut, -fon, -fik*, all starting with -f- to facilitate stem-affix segmentation), randomly allocated to cells in each paradigm for each participant such that both paradigms share the same number of unique forms but differ in their implicative structure (measured by i-complexity). In the low i-complexity paradigm, the singular form of a word predicts the dual form, while in the high i-complexity paradigm it does not. Figure 1 shows two example paradigms. Note that the distinct plural forms in each paradigm serve to distinguish the three classes of nouns; without distinct plural forms, the low i-complexity paradigm would have fewer classes than the high i-complexity paradigm.

	Singular	Dual	Plural
noun class 1	-fir	-fut	-fon
noun class 2	-fir	-fut	-fel
noun class 3	-fob	-fam	-fik

(a) low i-complexity paradigm

	Singular	Dual	Plural
noun class 1	-fir	-fut	-fon
noun class 2	-fir	-fam	-fel
noun class 3	-fob	-fut	-fik

(b) high i-complexity paradigm

Figure 1: Example paradigm for low i-complexity (a) and high i-complexity (b) languages. In this example low i-complexity paradigm, knowing that the noun in singular ends with *-fir* can, in principle, assist in predicting that the form in dual ends with *-fut*. This is not the case for the high i-complexity paradigm.

We also manipulate the number of nouns assigned to each class in the language, i.e., the size of the noun class or its type frequency. In the balanced-frequency condition, each noun class includes an equal number of stems (6 stems). In the skewed-frequency conditions, stems are assigned to noun-classes forming uneven noun class sizes, as described in Table 1. The token frequency of the stems (i.e., number of repetitions of the same stem inflected for a grammatical number that participants encounter in each block of training) is contingent on the type frequency of its noun class; stems in smaller noun classes appear with higher token frequency, in order to balance overall frequency of each noun class. An additional set of 9 stems and paired meanings was used to test generalization of the paradigm to novel nouns.

Table 1: Noun classes type (red) and token (blue) frequency per condition. Type frequency reflects the number of stems inflected according to the noun class. Token frequency reflects the number of occurrences of inflected forms in the noun class per grammatical number, per block of trials in the task. E.g. 6 X 3 indicates 6 stems in a given class, each occuring 3 times per block.

	Balanced	Skewed class 1	Skewed class 3
Noun class 1	6 X 3	9 X 2	3 X 6
Noun class 2	6 X 3	6 X 3	6 X 3
Noun class 3	6 X 3	3 X 6	9 X 2

Participants were trained on the artificial language using a staged learning procedure; learners are first trained on the singular forms of the nouns in the language, after which they are exposed to both singular and plural forms, and finally dual forms are included. The critical trials in our experiment are the dual items, since it is the predictability of the duals that differs across the low and high i-complexity paradigms. We test how well learners learn the dual forms in the language after being exposed to the singular and plural inflected forms, and how well they are able to generalize to the dual form of novel nouns when given that novel noun in the singular.

Participants 233 self-reported native English speakers participants were recruited via the Prolific crowd-sourcing platform. The mean duration of the task was 39 minutes and participants were compensated £6 for their time. Participants were allocated randomly to each of the six i-complexity/frequency paradigm type parinings: low-i/balanced (40); high-i/balanced (43); low-i/skewed-class 1 (36); high-i/skewed-class 1 (39); low-i/skewed-class 3 (38); high-i/skewed-class 3 (37).

Procedure The task consists of three parts: initial attention trials, learning the forms; generalization to novel stems.

During the attention trials (6 trials at the beginning of the task and an additional 6 trials randomly dispersed over the first block of learning trials), participants are presented with a picture of a simple object and are asked to choose the correct English name for this object from a set of possible labels. These trials are used to filter out innattentive participants.[1]

[1]No participant was filtered out for failing the attention trials.

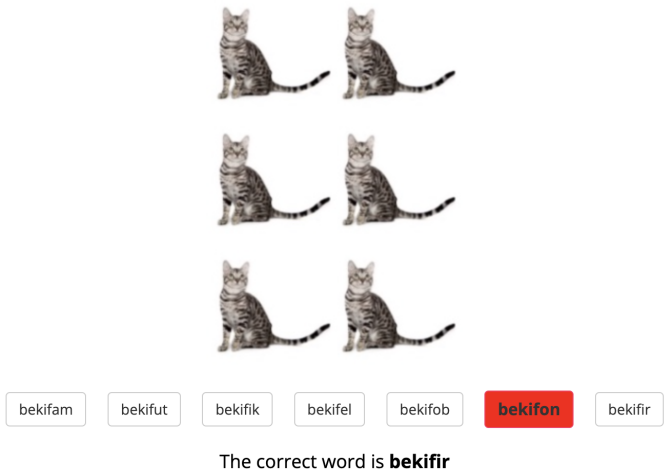

Figure 2: Screenshot of an example trial from the behavioural task. On this trial, the participant chose the incorrect label for cats (plural) in the artificail language and is presented with the correct word, *bekifir*.

On each trial in the learning phase, a picture is presented on the screen together with a set of possible stem + suffix labels (see Figure 2). All labels include the correct stem for the presented object together with each of the seven suffixes in the language. Participants are asked to choose the correct label, and receive feedback on their answer (whether they selected the correct or incorrect form, and in the case of incorrect responses, what the correct form was). The learning task is divided into 3 blocks of trials. In block 1 (54 trials), participants are exposed to the singular forms of all stems. In block 2 (108 trials) plural trials of all stems are introduced along with singulars. In the critical block 3 (162 trials), participants are exposed to all stems in all cells of the paradigm, including the dual. The number of presentations of each word form (token frequency) in each block differs between the balanced-frequency and skewed-frequency conditions, as described in Table 1. The different forms are randomly interspersed within each block.

The fourth and final block of trials forms the generalization phase of the task; in this block, participants are asked to choose the correct dual label for novel items. Generalization trials work in the same way as learnign trials (participants see an image and select an inflected form, with feedback) but feature novel nouns and occur in pairs: the participant is tested on the singular for a novel noun (allowing them to see the appropriate inflected form), and are then immediately tested on the dual form of the same noun, requiring them to attempt to generalise from the singular to the dual form. The generalization block consists of 18 trials (two successive trials of each of the 9 novel items). Since the target trials in our design are the dual forms, we focus on them when analysing the generalization data.

Hypotheses Our hypotheses for effects on learning and generalizing the forms in the paradigm are summarized in Table 2.

Based on prior work (e.g. Johnson et al., 2020; Seyfarth et al., 2014; Copot & Bonami, 2024) we expect that low i-complexity paradigms will be more rapidly or accurately learned than high i-complexity paradigms (hypothesis 1). We also expect (based on e.g. Gómez, 2002; Tenenbaum & Griffiths, 2001) that high type frequency (i.e., larger noun class displying higher stem variability) will facilitate generalization to novel nouns (hypothesis 2).

Under the Marginal Detraction Hypothesis, low type frequency inflectional classes are less dependent on the implicative structure of the paradigm and thus can conform less to its predictive structure. We therefore predict (hypoithesis 3) that the classes with low type frequency will be learned more accurately (class 1 is marginal in the skewed class 3 condition and vice versa). Note however that these same effects are also predicted as effects of token frequency: participants in the skewed class 1 condition are exposed to the three stems of noun class 3 more frequently during training (and the same for noun class 1 for the skewed class 3 condition, see Table 1) which may assist in learning those forms.

Another prediction from the Marginal Detraction Hypothesis that would not be explained by token frequency only, albeit a more strict prediction, is that the effect of i-complexity is moderated by type frequency. In other words (hypothesis 4), the effect of i-complexity will be smaller for marginal (low type frequency) inflectional classes, as they are learnable without conforming to the implicative structure.

Results

Learning the Forms Figure 3 shows the mean accuracy with which participants chose the correct label as the dual form for the presented object in the third block of the learning phase of the experiment. Participants' accuracy was on average higher than chance towards the end of the learning task, suggesting they were able to learn the inflected dual forms in the language.

We analysed this data using a mixed-effects logistic regression model predicting accuracy in dual trials by trial number (scaled), accuracy in block 2 (scaled)[2] i-complexity, type frequency, and noun class.[3] Contrasts for the type fre-

[2]Participant's accuracy in block 2, prior to when the dual trials are introduced and the conditions diverge, was added to the model as a way of controlling for general differences in learning ability between participants.

[3]We included noun class since it is predicted to interact with other fixed effects, as outlined in Table 2. However, the model also revealed a significant main effect of noun class 1 ($b = -0.52$, $z = -14.8$, $p<0.001$) and noun class 3 ($b = -0.4$, $z = -11.85$, $p<0.001$), indicating that participants labeled the dual forms in noun class 1 and noun class 3 with higher accuracy compared to forms inflected according to noun class 2. The analysis of generalization data also showed similar effects of noun class 1 ($b = -0.3$, $z = -3.5$, $p<0.001$) and noun class 3 ($b = -0.45$, $z = -5.2$, $p<0.001$), indicating lower performance on noun class 2 stems. This effect of noun class was not part of our hypotheses and we do not have a clear theoretical explanation for it.

Table 2: Summary of the hypotheses for our experimental design. Predictions are with respect to dual forms specifically.

	Hypothesis	Prediction
H1	I-Complexity effect	higher accuracy in low i-compelexity paradigm
H2	Type frequncy effect	higher generalization accuracy in noun class 1 in skewed-1 (and in noun class 3 in skewed-3)
H3	Marginal Detraction Hypothesis (also simple token frequency effect)	higher accuracy in noun class 1 in skewed-3 (and in noun class 3 in skewed-1)
H4	Marginal Detraction Hypothesis	Smaller effect of i-complexity on learning noun class 1 in skewed 3 (and noun class 3 in skewed-1)

quency fixed effect were set such that the balanced condition is the reference level and the two other conditions, skewed-1 and skewed-2, are compared to it. The model included by-participant intercepts and random slopes for trial number.

This model revealed a significant effect of i-complexity (b = -0.39, z = -5.74, $p<0.001$); participants exposed to the low i-complexity paradigm are better in learning the dual forms, as predicted in hypothesis 1.

There was a pattern of significant interactions suggesting evidence for the Marginal Detraction Hypothesis (hypothesis 3) (skewed-1 type frequency and noun class 1, b = -0.22, z = -4.52, $p<0.001$; skewed-3 type frequency, and noun class 1, b = 0.3, z = 5.81, $p<0.001$; skewed-1 type frequency and noun class 3, b = 0.28, z = 5.47, $p<0.001$; skewed-3 type frequency and noun class 3, b = -0.32, z = -6.64, $p<0.001$)[4]. As metioned above, these effects could also (and perhaps more simply) be explained through effects of token frequency.

Concerning the predicted interaction of i-complexity and type frequency stated in hypothesis 4, however, our regression model showed no significant three-way interaction between skewed-1 type frequency, i-complexity and noun class 3 ($p=0.34$) or between skewed-3 type frequency, i-complexity and noun class 1 ($p=0.98$).

Results from block 3 of the learning task suggest that, for familiar nouns encountered throughout blocks 1—3 of the experiment, a) the implicative structure of the paradigm (low i-complexity) facilitates learning the forms in the paradigm, b) high token frequency assists in learning inflected forms (note that since high token frequency corresponds to low type frequency in our design, this also suggests that high type frequency does *not* facilitate learning) and c) we see evidence for one learnability prediction generated by the Marginal Detraction Hypothesis, hypothesis 3, but this could be also explained by token frequency effects. We do not see evidence for hypothesis 4, that the effect of I-complexity is moderated by the size of the noun class, suggesting that the process by which marginal detraction can be explained closely relates to token frequency effects.

[4]Interactions between the above effects and trial number were also significant (b = -0.29, z = -3.36, $p<0.001$; b = 0.41, z = 4.58, $p<0.001$; b = 0.19, z = 2.2, $p<0.05$; b = -0.34, z = -4.07, $p<0.001$, respectively)

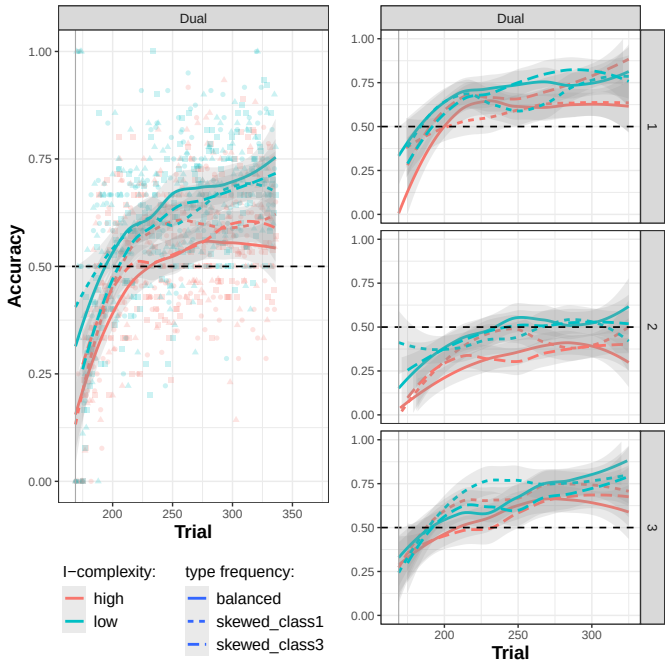

Figure 3: Mean accuracy by trial for all dual forms in the language (left) and for dual forms divided by noun class (right). Loess fit curves predicting accuracy by trial number for each of the conditions. Horizontal dashed line indicate chance level.

Generalizing to Novel Forms Figure 4 shows the mean accuracy with which participants generalized the dual forms for novel stems, after being presented with the form in singular. To test effects of i-complexity, type frequency and how they interact with noun class, a mixed-effects logistic regression model predicting accuracy in dual trials by i-complexity, type frequency, accuracy in block 2 (scaled) and noun class[5] was fitted to the data. The model revealed significant effect of i-complexity (b = -0.85, z = -10.6, $p<0.001$); as in learning, participants in the low i-complexity conditions were better at generalizing the dual forms to novel items.

Contrary to our prediction (hypothesis 2), there is no ev-

[5]The model also included by-participant intercepts. We did not include trial number since number of dual trials in the generalization task is small (9).

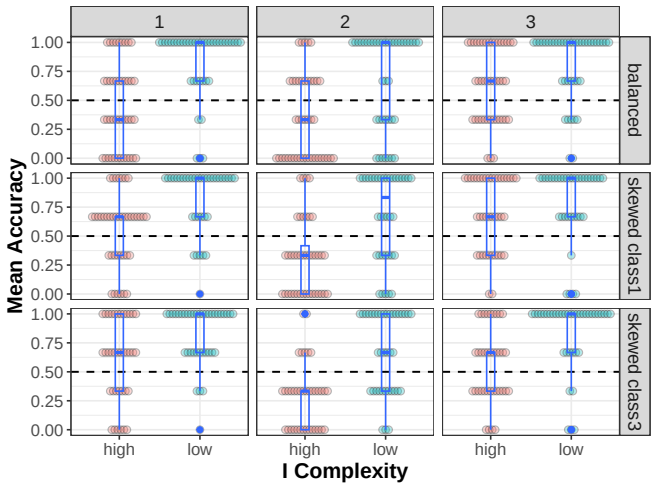

Figure 4: Mean accuracy in generalizing the dual forms to novel stems in the six type frequency, i-complexity conditions by noun class (column facets). Points in the figure represent participants' mean accuracy in generalizing the dual forms from each noun class.

idence for type frequency effects on generalization either in main effects of type frequency conditions(p=0.76 for skewed-1 type frequency and p=0.79 for skewed-3 type frequency) or in interactions between skewed-1 type frequency and noun class 1 (p=0.89) and skewed-3 type frequency and noun class 3 (0.42).

In respect to predictions from the marginal detraction hypothesis, the model showed no evidence for the three-way interaction between skewed-3 type frequency, i-complexity and noun class 1 (p=0.09) or skewed-1 type frequency, i-complexity and noun class 3 interaction (p=0.74). In other words, there is no evidence for effect of type frequency for the moderation of effects of i-complexity by type frequency when generalizing the duals to novel forms.

Simulations with ANNs

Next, we report results from simulations with artificial neural networks (ANNs) trained on data structured in a similar way to the artificial language presented to the human participants. Our motivation here is to use ANNs as surrogates for human learners in experiments with larger lexicons, which would be impractical to run with human participants. ANNs have been used as surrogates for human learners in the past (Johnson et al., 2021; Pimentel et al., 2021). We validate them here by comparing their results to those from our human experiment on similarly-sized lexicons, then run them on larger lexicons to discover which effects persist. To preview our findings: the results of type frequency mostly disappear in large lexicons; we discuss potential reasons why below.

Model Architecture

Following Johnson et al. (2021), we use 25-unit LSTMs with a single softmax classification layer following the last hidden unit. Thus, the LSTMs read their input (the stem of the word to inflect and a character indicating the desired form) character-by-character, but predict the affix as a unit, a choice motivated by the discrete and phonologically invariant forms of affixes used in this study. We use an SGD optimizer with a learning rate of 0.1.

To match the staged training of the human experiment, stimuli are presented to the network over 900 epochs divided into 3 stages of 300 epochs each. In stage 1, only singulars are presented; in stage 2, plurals are also presented; in stage 3, all forms are presented. During each epoch the entire set of forms is presented once in random order. We analyse our results by measuring the speed at which the network learns its training data, measured in epochs. Unlike a long series of prior ANN experiments focused on generalization behavior of majority and minority allomorphs (Hare, Elman, & Daugherty, 1995; McCurdy, Goldwater, & Lopez, 2020), we are concerned with how quickly the network memorizes its training set, which compares directly to our outcome measure during learning in the human experiment. The network experiences brief intervals of catastrophic forgetting at each new stage, but recovers quickly (e.g. the learning curve for singulars in stage 2 is much more rapid than in stage 1); we evaluate results only for the last 300 epochs.

Data

We run the network 50 times for each experimental condition, resampling the stems at random for each run, and assigning words to inflection classes at random. In addition to the conditions shown in 1, we simulate runs with 10 times more words in each class (large lexicons). For example, the large balanced condition has 60 word types per class for a total of 180 tokens. We chose to scale the lexicon multiplicatively because this preserves the proportional relationships central to our experimental design; however, it is not clear that these proportions have the same impact on learning as the lexicon scales. Sims and Parker (2016) treat marginality as a gradient relationship between type frequency and information-theoretic outcomes, so that any non-majority class is at least somewhat marginal, but the Tolerance Principle (Yang, 2016) or certain Bayesian models (Goldwater, Johnson, & Griffiths, 2005) would suggest a logarithmic rather than multiplicative threshold for marginality.

Results

We run similar regression analyses on the simulated data as on the experimental results. There are two differences: we check the network's accuracy on the entire dataset after each epoch and predict the log accuracies directly rather than using binary outcomes for single words. We drop certain random effects which lead to singular fits because the variance across network runs is very low compared to human participants.

The effects of noun class and I-complexity are comparable to those for humans. However, in the model we obtain main effects for the type frequency condition (for small languages only) rather than the complex pattern of interactions observed with human participants, and for large languages, we see no type frequency effects at all. We return to these issues in the discussion.

In small lexicons, we find a significant result of I-complexity (b=-.03, $p<0.001$), i.e. as for human learners, low i-complexity facilitates learning of dual forms. There was an interaction between skewed-3 type frequency and noun class 1 (b=-0.12, $p<0.01$), supporting the Marginal Detraction Hypothesis (hypothesis 3). The model however did not show an intercation between skewed-1 type frequency and noun class 3, also predicted in the same hypothesis. In respect to hypothesis 4, the regression model did not reveal significant three-way interactions between skewed-1 type frequency, i-comlpexity and noun class 3 (p=0.27) or skewed-3 type frequency, i-complexity and noun class 1 (p=0.12). As in the results from the human learners, there was no evidence for effects of the Marginal Detraction Hypothesis that are orthogonal to token frequency effects. The different type frequency conditions also had significant effects; skewed class 1 paradigms were easier than balanced frequency paradigms (b=-.014, $p<0.001$), while skewed class 3 paradigms were harder (b=.014, $p<0.001$).[6]

From running the LSTMs on a similar task as the human participants we see similar patterns of results in respect to effects of i-complexity (hypothesis 1), effects predicted by the Marginal detraction Hypothesis and token frequency (hypothesis 3, though only partially) and no independent effects of Marginal Detraction Hypothesis (hypothesis 4). While results from the behavioural experiment did not show effects of type frequency on learning or generalization of the forms, results from the LSTMs suggest that learning was affected by type frequency.

For large lexicons, the effects of I-complexity and noun class remain significant and in the same direction as for small lexicons, while type frequency does not.

Discussion

Our results show that the implicative structure of the paradigm affects both learning the inflected forms (in humans and the artificial neural networks) and generalizing the paradigm to novel words, while in Johnson et al. (2021, 2020), the implicative structure was found to have a weaker effect, secondary to the number of different inflection endings. This could be due to differences in the design of the behavioural task with the human participants that made the implicative structure more apparant in our study (here we did not manipulate number of distinct endings; all suffixes started with *f* to help segmentation and order of buttons array of the label endings was kept constant on the screen over all trials which may have facilitated the task in general). Another possibility is that the effect of the implicative structure with human learners is not very stable and therefore comes out significant in some cases and not in others. Calculating the effect size in a meta analysis of the published studies could help in better understanding the role of implicative structure in language learning.

We also show that token frequency assists in learning the forms in the paradigm in human learners. While prior studies (Copot & Bonami, 2024; Hare et al., 1995) tend to focus on the role of type frequency rather than token frequency, these are typically studies of morphological generalization— the role played by token frequency (repetition) in learning specific lexical items from experience is well known. In generalization, we did not see effects of either type or token frequency. Our prediction was that type frequency (and not token frequency) would facilitate generalization to novel forms; it could be that much larger differences in type frequencies are necessary to obtain such an effect.

The neural networks show several of the same effects as human learners, except for type frequency, where we find main effects rather than interactions. This suggests that LSTMs remain reasonable, though not perfect proxies for humans in this task setting. Testing the ANNs with larger-scale languages (which cannot be taught to humans in the lab) indicated no effects of type frequency at all, except in a single interaction with noun class 3, which suggests that frequency effects may not be stable as lexicon sizes scale up. This could be because real irregular classes generally have logarithmically many tokens compared to regular ones (e.g. Yang, 2016), while we scaled each class size by a constant multiple; therefore, the large-scale setting may no longer be an effective model of irregular or marginal inflection. Future experiments could investigate whether marginal detraction effects apply to all minority classes in the lexicon, as implied by the analysis of Sims and Parker (2016), or occur only for sublogarithmic or even smaller classes which can be treated as truly exceptional. We remain interested in using neural models to investigate the potential outcomes of artificial language learning experiments too large to conduct with real participants, in order to compensate for the artificiality of language learning tasks with only a few dozen items.

Overall, our results support predictions made by the Marginal Detraction Theory (Stump & Finkel, 2013; Sims & Parker, 2016) although they are limited to cases where the hypothesis can be also explained by token frequency effects. The implicative structure of inflectional paradigms was shown to be a factor that needs to be taken into account in studies looking at morphological complexity in terms of language learning and processing.

Acknowledgments

This research also received funding from the European Research Council (ERC) under the European Union's Hori-

[6]The model also revealed a significant effect of noun class. Noun class 2 and 3 were both harder to learn than noun class 1 (b=0.027, 0.071, $p<0.001$).

zon 2020 research and innovation program (Grant Agreement 681942), held by Kenny Smith.

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 429–464.

Ackerman, F., & Malouf, R. (2015). The no blur principle effects as an emergent property of language systems. In *Proceedings of the annual meeting of the berkeley linguistics society* (Vol. 41).

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425–455.

Copot, M., & Bonami, O. (2024). Baseless derivation: the behavioural reality of derivational paradigms. *Cognitive Linguistics*(0).

Cotterell, R., Kirov, C., Hulden, M., & Eisner, J. (2019). On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, *7*, 327–342.

Goldwater, S., Johnson, M., & Griffiths, T. (2005). Interpolating between types and tokens by estimating power-law generators. *Advances in neural information processing systems*, *18*.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436. Retrieved from https://doi.org/10.1111/1467-9280.00476 (PMID: 12219809) doi: 10.1111/1467-9280.00476

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and cognitive processes*, *10*(6), 601–630.

Johnson, T., Culbertson, J., Rabagliati, H., & Smith, K. (2020). *Assessing integrative complexity as a predictor of morphological learning using neural networks and artificial language learning.* PsyArXiv. Retrieved from osf.io/preprints/psyarxiv/yngw9 doi: 10.31234/osf.io/yngw9

Johnson, T., Gao, K., Smith, K., Rabagliati, H., & Culbertson, J. (2021). Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems. *Journal of Language Modelling*, *9*(1), 97–150.

McCurdy, K., Goldwater, S., & Lopez, A. (2020). Inflecting when there's no majority: Limitations of encoder-decoder neural networks as cognitive models for german plurals. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 1745–1756).

Parker, J., & Sims, A. D. (2020). Irregularity, paradigmatic layers, and the complexity of inflection class systems: A study of russian nouns. *The complexities of morphology*, 23–51.

Pimentel, T., Leonard, B., Ryskina, M., Mielke, S., Haley, C., Chodroff, E., ... Ambridge, B. (2021). *Are we there yet? A shared task on cognitively plausible morphological inflection* (Tech. Rep.). Retrieved from https://github.com/sigmorphon/2021Task0

Seyfarth, S., Ackerman, F., & Malouf, R. (2014). Implicative organization facilitates morphological learning. In *Annual meeting of the berkeley linguistics society* (Vol. 40, pp. 480–494).

Sims, A. D., & Parker, J. (2016). How inflection class systems work: On the informativity of implicative structure. *Word Structure*, *9*(2), 215–239.

Sims-Williams, H. (2016). Analogical levelling and optimisation: The treatment of pointless lexical allomorphy in greek. *Transactions of the Philological Society*, *114*(3), 315–338.

Stump, G., & Finkel, R. A. (2013). *Morphological typology: From word to paradigm* (Vol. 138). Cambridge University Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*(4), 629–640.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.