Is there scope for scope in morphophonological rule induction?

Núria Bosch and Bert Vaux
University of Cambridge

1 Introduction: (some) evaluation metrics in rule induction

In the context of finite input, learners are faced with a learnability problem: they must infer the full set contexts where a given rule or dependency applies, even if the input exposed contains only a limited set of tokens relevant to that given rule. In order to make headway in situations of input sparseness, the learner is thus forced to extrapolate. A key question in language acquisition, both phonological and morphosyntactic, remains which 'extrapolation' or, more precisely, *evaluation metrics* the learner uses in the task of inferring one linguistic system over other possible systems (Gold, 1967).

An oft-cited 'dichotomy'[1] between possible metrics is one between a *conservative* bias and a *simplicity* bias. A conservative bias drives the learner to be maximally faithful to the input, and not to deviate or extrapolate beyond it: the Subset Principle (Berwick, 1985), exemplar-based learning (Johnson, 1997; Pierrehumbert, 2001) and some implementations of Bayesian models (e.g., Tenenbaum, 1999) are examples. A simplicity bias, on the other hand, impels the learner to favor formally simple(r) rules. 'Formally or notationally simple(r)' can be construed in a range of ways, but a common interpretation is in terms of feature economy or minimization, as in, i.a., Halle (1961) and Chomsky & Halle (1968). That is, all other things being equal, a simplicity-biased learner should pick r-insertion rule (1a) over (1b). We return to (1) later below.

(1)
a. $\text{Ø} \rightarrow$ /r/ / $V_{[\text{-hi}]}$ _ V, **over**
b. $\text{Ø} \rightarrow$ /r/ / $V_{[\text{-hi, -rd, +bk}]}$ _ V

A range of evidence has been gathered for both simplicity and/or conservative biases within and beyond phonology: computational work (i.a., Gold, 1967; Albright & Hayes, 2002; Carr et al., 2020), synchronic and diachronic work (i.a., Chomsky, 1957; Fodor & Crain, 1987; Clark & Roberts, 1993) as well as experimental work (i.a., Pycha et al., 2003; White, 2013; Culbertson & Kirby, 2016).

In this paper, we propose to move beyond this traditional dichotomy and explore the possibility of learning biases that hinge on *rule scope* (Vaux, 2008). More specifically, Nie et al. (2019) discuss the plausibility of a *scope expansion* bias, which prioritizes rules that apply to as many segments as possible and *independently* of the notational simplicity of the resulting rule (Vaux, 2008). Scope expansion raises the possibility of a bias which often, but *not always,* overlaps with simplicity. Elaborating on (1) above, expansion in r-insertion in Southern Standard British English (SSBE) is one such case study where simplicity and scope expansion equally successfully account for the data: r-insertion occurs in four distinct stages (Wells, 1982), fleshed out in (2). Given standard metrics of featural simplicity, we can observe that they are usually highly analogous: later stages result in successive deletion of features. It follows, too, that the rule

* Our thanks go to Samuel Andersson, for very useful help and suggestions, and to Theresa Biberauer and Ollie Sayeed, for discussion. Thank you also to attendees and reviewers of AMP 2024 for their useful feedback. The first author is generously supported by an Open-Oxford-Cambridge AHRC DTP – St John's Studentship, co-funded by AHRC (UKRI) and St John's College.

[1] In scare quotes. Despite the often-opposed nature of these two biases, it remains a possibility that both simplicity and conservativeness biases co-exist in complementary ways. An example of this could be found in theoretical syntax, where co-existing biases of Feature Economy (minimization of features) and Input Generalization (maximal use of available features) have been proposed (see Roberts & Roussou, 2003; Roberts, 2007/2021; and subsequent work).

Proceedings of *AMP 2024*

broadens in scope, applying to more vocalic contexts.

(2) Expansion of contexts triggering r-insertion in SSBE (from Nie, 2016:4)
  a. Stage 1: /ə/, including centering diphthongs
     Ø → [-high, -low, +back, -round, +ATR] __]σ V
  b. Stage 2: /ə ɑː/
     Ø → [-high, +back, -round] __]σ V
  c. Stage 3: /ə ɑː ɔː/
     Ø → [-high, +back] __]σ V
  d. Stage 4: /ə ɑː ɔː aʊ (u)/
     Ø → [+back] __]σ V

In contrast, Ross (2011) and Nie et al. (2019:28ff) report different trends in scope expansion in the inherited /o/-lowering in Schaffhausen dialects, as illustrated in (3). In all dialects, /o/ lowers to [ɔ] before /r/ (3a). In some dialects, /o/-lowering was extended to all nasals and /r/ (3b). (3b) instantiates another case of featural simplification leading to scope expansion: [-nasal] is deleted in (3b) from the diachronically-original conditioning environment (3a). However, in neighboring villages, scope expansion was also favored, but *not* in ways that ultimately lead to simplification. /o/-lowering extended from applying before only /r/ to nasals and coronal obstruents as well (3c). The change was even more extensive in 5 of these Schaffhausen dialects, where /o/-lowering now operates before all obstruent except /b/ (3d). Now note how scope expansion favors (3c) and (3d) over (3a-b), appropriately aligning with the observed diachronic pattern. A simplicity bias does not; (3b) (or 3a) should prevail, by this metric.

(3)
  a. /o/ → [ɔ] before /r/
     [-high, +back] → [-ATR] / __ [+cons, +son, -lat, -nasal]
  b. /o/ → [ɔ] before /r/, nasals (featurally simplest)
     [-high, +back] → [-ATR] / __ [+cons, +son, -lat]
  c. /o/ → [ɔ] before /r/, nasals, coronal obstruents (featurally more complex)
     [-high, +back] → [-ATR] / __ [+cons, +son, -lat] ∨ [+cons, +cor, -lat, -nas]
  d. /o/ → [ɔ] before all obstruents except /b/ (featurally more complex)
     [-high, +back] → [-ATR] / __ [+cons, +son, -lat] – [+lab, +voi, -nas]

Here, simplicity and scope expansion diverge: all other things equal, simplicity should not have favored some of the synchronically observed patterns in these dialects of German. Scope expansion crucially does, raising the possibility that scope expansion, not just simplicity, is at play in learner evaluation metric(s).

This underlines the necessity for a deeper investigation into finer-grained differences across learning biases, notably on scope expansion, which remains comparatively understudied. Diachronic data on rule change appears to support its existence (see Nie et al., 2019), at least to some extent. Some computational modelling (Sayeed & Vaux, 2023) has also incorporated scope expansion, along with simplicity biases, to maximize model fit. However, there is no work on scope expansion (or contraction) in real-time language learning and acquisition more broadly. This is our contribution in this paper: we present a *first attempt* at probing the effect of rule scope in an Artificial Language Learning (ALL) scenario[2]. This work is thus not

[2] An anonymous reviewer notes that the link we ostensibly assume above between the diachronic trends observed and their putative acquisitional origin is non-trivial (see, e.g., much work on Evolutionary Phonology; Blevins, 2004). We do not have space to go into the connections between diachronic morphophonology and acquisition. However, we note that the trend of scope expansion and overgeneralization has been noted in *both* phonological acquisition and change (Nie, 2016, for discussion and references). We adhere to generative traditions that take acquisition to be a primary locus of change (which follow Paul, 1880; Müller, 1890, i.a.; e.g., Hale, 2007). If this position is adopted, then it immediately follows that (at least some) diachronic trends conceivably reflect acquisitional biases. This makes it worthwhile to probe the extent to which diachronic trends find parallels in acquisition (see, e.g., Cournane, 2017, 2019; Biberauer, 2019, for further justification), and that is the spirit of our paper.

intended to be conclusive regarding these new questions, but rather to serve as a springboard for more in-depth, future research.

We now lay out our research questions. Previous work on scope expansion (e.g., Nie et al., 2019) typically considers one main way in which this bias may manifest, namely in the *output* of acquisition, such as via contextually broader rules (as in 3 above). In this paper, importantly, we incorporate a *two-way* perspective on how rule scope may bias acquisition: our primary question concerns its putative role as an *independent learning bias*. This is our first research question, namely whether scope expansion is observed as a learning bias, including, crucially, in contexts where a formally more complex rule is prioritized over featurally simpler alternatives. We formulate this more broadly as in RQ1:

> **RQ1 (Output-oriented).** Which generalization biases do participants exhibit when presented with sparse input in an ALL setting?

The potential influence of rule scope in generalization patterns is importantly independent of its status as a separate evaluation metric. We also probe here the extent to which linguistic generalizations induced from sparse *input* in ALL contexts differ *as a function of the original scope of the rule presented in training*. This leads to an additional research question, which concerns the extent to which the 'scope' of the linguistic patterns presented originally is more or less likely to affect the outcomes of learners' evaluation metrics. We define this question below as RQ2:

> **RQ2 (Input-oriented).** Do participants exhibit different generalization patterns depending on the scope of the rule they are exposed to?

To address RQ1-2 our ALL set-up probes both the existence of generalizations supporting scope expansion (output), but also differences in generalization patterns as a result of the rule participants learn initially (input). We do so, especially the latter point, by incorporating a novel variable in the design – three Conditions at training, each of which presents evidence for an alternation with narrower or wider scope.

This paper is organized as follows: Section 2 describes the design and stimuli of the ALL experiment, and the predictions made by existing evaluation metrics. Section 3 describes the results: our findings provide evidence against conservative learners; high overgeneralization rates are observed, irrespective of condition. While we find no effect of scope, we argue that this is predicted by Yang's (2002, *et seq.*) Tolerance Principle: learners systematically induced elsewhere rules and overgeneralized the form with the highest type frequency. We discuss the ways in which learning strategy (explicit/implicit) sheds light on the findings obtained. Section 4 concludes and considers future directions for this work.

2 Methodology

2.1. *Participants and procedure* 76 adult native English speakers were recruited through Prolific. All participants had normal hearing and were not reported to have any known visual or language impairments. The experiment was conducted online on Gorilla (www.gorilla.sc; Anwyl-Irvine et al., 2020) on the participants' own computers. Instructions were presented before each experimental phase. The entire procedure took around 20-30 minutes to complete.

2.2. *Design* Participants were exposed to singular-plural pairs in an artificial language. Our design follows the Poverty of the Stimulus (PoS) paradigm in ALL (Wilson, 2003): participants initially learn only a subset of the artificial language, thereby missing evidence that would disambiguate between multiple possible hypotheses. After this initial learning phase, participants are then presented with testing stimuli which differ from exposure stimuli in critical ways. This unseen data helps discern which kinds of grammatical rules participants most readily adduce.

In this experiment, we presented participants with CVC pseudo-words with auditory evidence for an alternation in the plural suffix; some, but not all, plural suffixes underwent diphthongization, e.g., a contrast between /kap/ > /kap**wok**/ and /pen/ > /pen**ok**/[3]. All phonemes in the language were shared with English. Participants were explicitly asked to focus on the use of the plural suffix only, and to *ignore* the words' potential semantic denotation and the associated photos provided, which were randomly allocated.

The primary novelty in our PoS design rests in the incorporation of *rule scope* in the design. We designed three separate conditions manipulating the scope of the pluralization rule, to study the potential two-way effect of scope in inductive learning (§1). In the first *familiarization phase,* each condition provided positive evidence for diphthongization as follows: (i) the Plosives (or intermediate scope) Condition showed diphthongization after stems ending in a plosive; (ii) the Voiced Plosives (narrow scope) Condition showed diphthongization after *voiced* plosives only; (iii) the Obstruents (wide scope) Condition displayed evidence for it after all obstruents. The three Conditions are in 'containment' relationships with one another – from narrower to wider: Voiced Plosives > Plosives > Obstruents. Participants observed negative evidence for diphthongization in nasals and approximants, for the wider conditions (Plosives and Obstruents), and for nasals only in the Voiced Plosives Condition[4]. As a 'control', we *held out* one phoneme per natural class: these were /b/, /t/ (plosives), /z/, /v/ (fricatives), /r/ (approximant) and /ŋ/ (nasal). These held-out environments controlled for potential rote memorization of the training stimuli, and so help establish the likelihood that participants extracted more abstract morphophonological generalizations. Participants had to listen to all stimuli in this task, both singular and plural. The morphophonological rules which we envisaged as underlying the three conditions are stated in (4)[5]:

(4)
a. Plosives Condition: /ok/ → /wok/ / [+cons, -son, -cont] __
b. Voiced Plosives Condition: /ok/ → /wok/ / [+cons, -son, -cont, +voi] __
c. Obstruents Condition: /ok/ → /wok/ / [+cons, +son, -lat] __

There were 23 participants in the Plosives Condition, 25 participants in the Voiced Plosives Condition and 28 in the Obstruents Condition.

The second task, the *evaluation phase*, incorporated a quiz measuring the learning success of the patterns in the exposed trials. This was assessed using a forced-choice task where participants had to choose between plural forms (suffixed with either /ok/ or /wok/) of already-seen singular stems. Participants with a learning score below 70% were excluded from the analysis.

The third, *testing phase* was also a forced-choice task. Following the PoS paradigm, the aim was to study the inductive trends observed when participants are forced to extrapolate to novel contexts. Participants were asked to infer plural suffixes (/ok/ or /wok) for 54 entirely new stems, for both seen and unseen consonantal environments, based on their knowledge of the artificial language from training. This phase was identical across conditions. A sample screen for the forced-choice task is given below:

[3] A reviewer asks why plural-suffix diphthongization was chosen, given it is not a typologically widespread phenomenon. The choice of diphthongization and pluralization was made to facilitate learning and perception of the alternation by English native speakers. We also did so to avoid potential confounds with naturalness (i.e., phonetically more natural alternations); we think no such naturalness factors could be hypothesized by participants for the alternations in the study. This aside, it is also worth remarking that if phonological computation is in any sense 'substance-free' (Hale & Reiss, 2008), the naturalness or typologically commonality of this rule are orthogonal to our aim; the generalization patterns observed with this rule are arguably as consequential as other conceivable rules. If anything, the choice of a rarer phenomenon avoids additional variables in a welcome way, such as phonetic naturalness, that do arise with other phenomena (e.g., consider the evolution of umlaut plurals in several languages).
[4] To ensure a balanced distribution of (non)diphthongization across types; see §2.3.
[5] These reflect our rationale in the experiment design. However, we emphasize that these rules are only some of many logically possible rules given the input participants are presented with, as is standard in PoS settings. We return to this at various points.

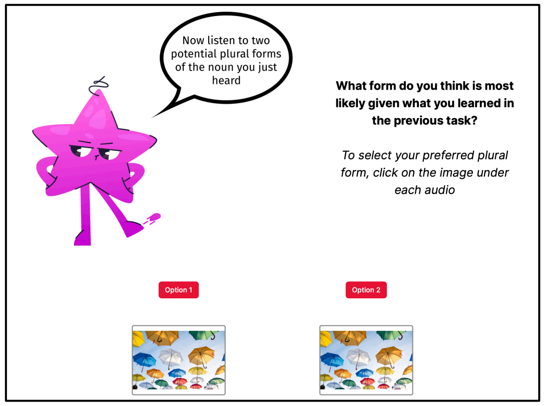


Figure 1. Sample screen for testing task.

Upon completion of the study, participants were directed to a short debrief section, where they reported their learning strategies, a forced-choice between ‘rule-based’ *vs* ‘intuition-based’.

Finally, to proceed with the data analysis, we calculated several scores from the data set obtained: a learning success score (% correct) in the evaluation quiz, an input-faithfulness score during testing (estimating participants’ ability to extend existing knowledge to *new* stems with familiar consonantal endings). We also coded for overgeneralization and undergeneralization (calculating how much more or less diphthongization was observed relative to the input presented). Undergeneralization was in practice rarely observed, however (82% of instances of input deviation were overgeneralizations). The variable is dropped from the statistical analyses and discussion in §3, as it did not explain more variance than the overgeneralization variable alone[6].

Participants with an input faithfulness score lower than 65% were discarded in statistical analyses of the testing phase, insofar as participants had not extracted patterns that generalized to novel stems with familiar conditioning contexts[7]. Attention checks were interspersed across all phases of the experiment. Participants who failed one or more attention checks were also excluded. This led to a total 47 participants for the familiarization and evaluation phases, and 33 participants after the testing phase.

2.3. *Stimuli* All stimuli were produced synthetically via a speech synthesizer. Audio and other (e.g., visual) stimuli were screened in a pilot round for intelligibility and for errors.

The familiarization phase contained three sets singular/plural (stem *vs* stem+suffix) pairs for each Condition as described above: 40 items (20 stems, 2 stems per consonant; each repeated twice) for the Plosives Condition, 32 items (16 stems, 4 stems per C; each repeated twice) for the Voiced Plosives Condition, and 60 items for the Obstruents Condition (30 stems, 3 stems per obstruent, 6 per nasal/approximant; each repeated twice). No statistical bias towards diphthongized forms (or otherwise) was present in the experimental design: the token frequency was counterbalanced, so that the items triggering diphthongization or not (e.g., obstruents *vs* nasals/approximants in the third condition) were (approximately) 50% each.

Examples of the paired items for the three Conditions and (some) environments can be found in Table 1 below. The training items were in containment relationships across Conditions, such that those in the Voiced Plosives, Plosives, and Obstruents Conditions were subsets of each other respectively.

[6] This, on its own, is already significant: the same trend is observed in child language acquisition errors, which often occur in the direction of expansion/overgeneralization (e.g., Stemberger & Bernhardt, 1997).

[7] Note that the overgeneralization score and the input-faithfulness score are necessarily inversely related, and, on these grounds, an appropriate criterion of exclusion based on faithfulness must be followed, to avoid the exclusion of participants who may be highly overgeneralizing despite having acquired some morphophonological rule. To reduce arbitrariness in the exclusion threshold, we used *k*-means clustering to divide the participant sample; this resulted in the 65% threshold used above for the main statistical analysis. Importantly, though, we reconsider the data of highly unfaithful participants in §3.2, when we discuss idiosyncratic, more localized behaviors.

Condition	**Plosives**	**Fricatives**	**Nasals**	**Approximant**
Plosives	/kap/ – /kapwok/ /rag/ – /ragwok/		/tem/ – /temok/ /zon/ – /zonok/	/dal/– /dalok/
Voiced Plosives	/kap – /kapwok/ /mek – /mekwok/		/tem/ – /temok/ /zon/ – /zonok/	
Obstruents	/kap/ – /kapwok/ /rag/ – /ragwok/	/bef/ – /befwok/ /vaʃ/ – /vaʃwok/	/tem/ – /temok/ /zon/ – /zonok/	/dal/ – /dalok/

Table 1. Examples of stimuli during familiarization.

The evaluation quiz contained the same stimuli as in the previous phase, the only difference being that each stem was only presented once (not twice, as in the familiarization phase). This therefore meant 20, 16 and 30 stimuli were used for each condition, respectively.

The testing phase contained items ending in all possible word-final consonants in English, including seen and unseen environments. The test set of stimuli was identical for all participants in all conditions. Three separate new stems were designed for every consonant (18 consonants in total); 54 items in total were presented in the force-choice task. All items for both familiarization and test sets appeared in a different random order for each participant. Each item (both the plural and singular form) was assigned to a photo visualizing their denotation (singular *vs* plural objects/entities); again, this pairing was randomly allocated. Examples can be found below:

	Plosives	**Fricatives**	**Nasals**	**Approximants**
All Conditions	/nap/ – /napwok/ /zot/ – /zotwok/ /gid/ – /gidwok/ /lig/ – /ligwok/	/tuf/ - /tufwok/ /diθ/ – /diθwok/ /buz/ - /buzwok/ /seʒ/ - /seʒwok/	/bom/ – /bomok/ /ken/ – /kenok/ /saŋ/ – /saŋok/	/ʒil/ – /ʒilok/ /θer/ – /θerok/

Table 2. Examples of test stimuli.

2.4. *Predictions* We now lay out our predictions for the generalization patterns in pluralization learning. As for RQ1 (§1), if learners are biased by conservativeness, we expect a *tight fit* to the input data, with little to no overgeneralization or deviation from the items presented. As such, we would not expect extension of diphthongization beyond the morphophonological environments at familiarization (e.g., Plosives for the first condition). If they are biased by some kind of simplicity preference, we anticipate a rule with maximal formal simplicity to be favored. Such a rule could take multiple forms: for example, it could lead to minimization of formal features (e.g., deletion of a [+voice] feature in Condition 2), elsewhere-rule formation and exception-based rules, among several others. Finally, if the learner exhibits at least some evidence for a scope bias, then extraction of a rule with maximal segmental applicability should be preferred, and this rule need not result in a featurally-simpler outcome (e.g., it could feature union/disjunction of two natural classes, as observed in 3c earlier).

There is a second way in which we could envisage rule scope having an effect, as pointed out by RQ2. Work probing RQ2 is absent, to the best of our knowledge. We largely leave predictions for RQ2 open, but we can conceive of some possibilities: for example, when presented with a very narrow rule (condition 2), participants could be more likely to abandon the alternation altogether, or would find it harder to acquire. Conversely, a much wider rule could facilitate overgeneralization and extension to novel contexts. At any rate, whichever patterns are observed as a function of scope in the input in our experiment, these will be important explananda.

3 Results and discussion

We first report *global* patterns observed in most of the data obtained (§3.1), supplemented with statistical analysis, before discussing other significant, but more *localized*, individual, patterns (§3.2). We offer an overall discussion of the results in §3.3.

3.1. *Global results* We introduce three sets of consequential global patterns: (i) overgeneralization of

/wok/ across all conditions, (ii) type, not token, frequency conditioning generalizations and (iii) predominance of explicit learning strategies in the task. In doing so, we report the results of a mixed effects logistic regression model with *lme4* in R (Bates et al., 2015; R Core Team, 2024).

First, we observed clear positive rates of overgeneralization of /wok/, over /ok/, across all three conditions. Crucially, the results showed *no effect of Condition* on overgeneralization rates in a one-way ANOVA ($F(2, 30) = 2.84$, $p = .074$). All three conditions thus showed comparable rates of overgeneralization, with the biggest difference being observed between the narrowest condition (Voiced Plosives) and the widest condition (Obstruents). This broad result is shown in Figure 2.

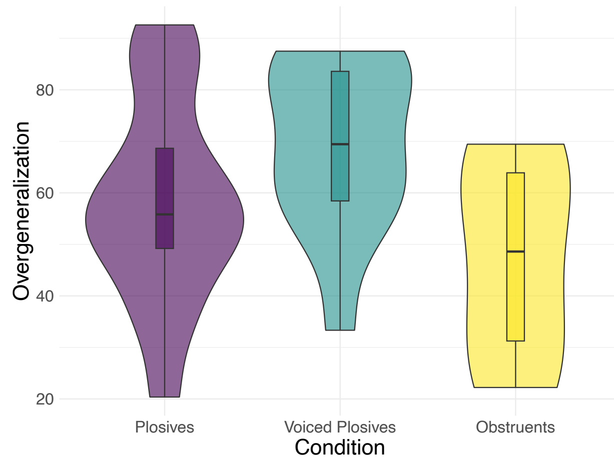


Figure 2. Overgeneralization rates by Condition.

Our mixed-effects logistic regression model confirms these observations in further depth. We analyzed the choice of plural suffix (/ok/ *vs* /wok/) in the testing phase as the dependent variable. Three independent variables (fixed factors) were included, namely the Condition, whether the consonantal environment was in the training data, the phonological natural class, as well as an interaction term of Condition*class. Participant ID was added as a random effect, giving the formula `choice ~ (1 | id) + Condition + intrain + class + Condition*class`. This was the maximal best-fit model (per BIC score) justified by the design of the experiment that would reach convergence, obtained with *glmulti*.

We observed that *all* segments and phonemic contexts are highly associated with *diphthongization* at testing ($p < .001$ for fricatives and plosives, relative to nasals as the reference level). This importantly also includes the *held-out* segments in each condition, e.g., /r/ and /l/, with a diphthongization rate of 77.78% and 75.55%, respectively. Only nasals and approximants, specifically the *individual segments* in the familiarization phase, were significantly less likely to trigger diphthongization ($\beta = -0.13$, $p = 0.610$), though with substantial variance. Segments outside of the training data were much more likely to trigger diphthongization, irrespective of phonological natural class ($\beta = 1.08$, $p < .001$). Figure 2 visualizes this.

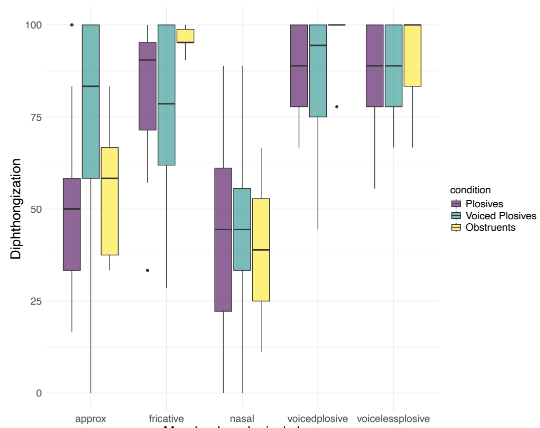


Figure 3. Percentage of /-wok/ by morphophonological class and condition.

Our hypothesis is that these results reflect *exception-driven learning* and elsewhere-rule formation: participants did not extract rules over morphophonological natural classes, but rather selected /wok/ as a 'default' or 'elsewhere' context, with (non-held-out) nasals and/or approximants being stored as exceptions; e.g., [PL] → /ok/ / /n, m, l/ ___; /wok/ ELSEWHERE. This result raises a range of issues for approaches assuming fundamentally conservative learners (e.g., Berwick, 1985; Snyder, 2007): learners did not limit the domain of diphthongization strictly to the segmental environments during familiarization. Instead, they are maximally 'generous' in its application to unseen stimuli, in a sense that we make more precise now.

An important question is why /wok/ is selected as the default over /ok/, if their (token) frequency was balanced in training. We now propose that the preference for /wok/ provides novel support for the salience of *type* frequency over *token* frequency in linguistic generalizations. This leads us to the **second** important result: we show, harnessing Yang's (2002, *et seq.*) Tolerance Principle (TP), that *type* frequency predicts the generalization patterns, including the lack of effect of Condition in Figure 2[8]. Incidentally, then, we provide a small piece of evidence that the TP, generally characterizing *child* acquirers, could be visible in *adult* learners too (cf. Schuler, 2017).

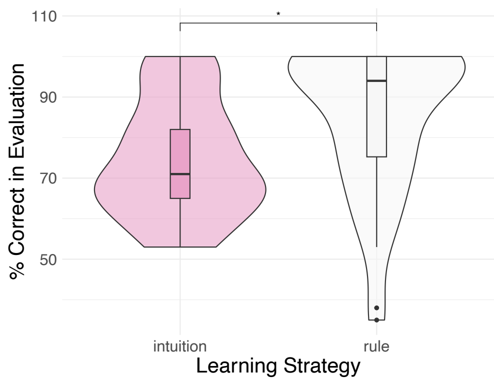
The average token frequency of diphthongized compared to non-diphthongized suffixes across the three conditions was 49-51% (47-53%, 50-50% and 50-50%, respectively) in the familiarization phase. There was less clearly a balance in type frequency, on the other hand. The Plosives condition contained 4 diphthongized types vs 3 non-diphthongized types (57-43%). Harnessing Yang's TP, we reach a calculation of $\theta_N = 4.59$. The Voiced Plosives condition had 2 types for each (50-50%; $\theta_N = 2.89$) and the Obstruent condition had 9 diphthongized types *vs* 3 non-diphthongized types (75-25%; $\theta_N = 4.8$). Per the TP, it follows that a productive rule can be formed in *all* conditions, since the 'exceptions' to the diphthongization trend fall always below the threshold specified by θ_N: the diphthongized suffix, by virtue of its higher type frequency, is chosen as the default suffix, with the rest of the contexts without diphthongization being memorized as exceptions to the rule. Note, the TP crucially explains the lack of significant effect of scope across conditions: the TP operates categorically, such that the relevant consideration is *only* whether a productive rule can be extracted relative to the relevant threshold. Insofar as all conditions meet this requirement (all exceptions being lower than θ_N), then we expect all conditions to behave comparably; degrees of 'closeness' to θ_N are irrelevant, and so should have no effect. It also explains why /wok/ is chosen over /ok/ as the primary plural suffix, across all participants; the former has higher type frequency and is chosen as the 'base' of TP's rule.

Importantly, we argue this supports a range of work showing that learners, especially children, are sensitive to type frequency and that they condition (some of) their extrapolations on this basis (i.a., Yang, 2002; Hay et al., 2004; Daland et al., 2011; Endress & Hauser, 2011; Schuler, 2017)[9]. Furthermore, it provides a case-study of adults tracking type frequency in generalizations, nuancing the claim that adults probability match in many cases (Hudson-Kam & Newport, 2005, and subsequent work). Yang's Tolerance Principle notably provides a unifying account of both the exception-driven learning noted earlier *and* the preference for /wok/ over /ok/ in these adult participants. These results underscore the need for more work on when and how type frequency steers adult learning and under which conditions it overlaps, or not, with TPs' predictions (see also Baer-Henney et al., 2015); cf. Schuler (2017), who argues that the TP characterizes children only.

A **third**, final element explaining the patterns observed in the ALL experiment is the role of learning strategies (e.g., implicit *vs* explicit learning). Explicit learning predominated in our experiment. Rule-based, explicit-learning strategies facilitated learning success (per self-reports, which are argued to be generally well-correlated with implicit/explicit learning; Merton & Pertsova, 2023): 'rule-based' participants performed significantly better in the pre-testing evaluation quiz than 'intuition-based' participants ($W = 249$, $p = .018$; Figure 3). They also showed higher faithfulness to the training input during the testing phase ($W = 95.5$, $p = .005$; Figure 4). Predominance of explicit learning reinforces the plausibility that exception-driven learning is at play in our data (see Yu et al., 2024).

[8] Note that it remains a possibility that the lack of significant effect is due to the sample size and the tests' predictive power. We set this aside and tentatively assume, following the only evidence we have, that there was really no effect of Condition. This should be verified in future work.

[9] This is directly reflected in the self-reports given in the debrief section, see one example from a participant: "I think that if the word ended with either l, m, or n, the plural ended in 'ock' rather than 'wok'".



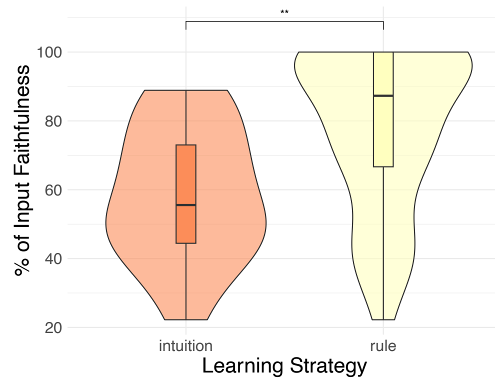


Figure 4. Performance in evaluation quiz by self-reported learning strategy | **Figure 5.** Input faithfulness by self-reported learning strategy.

These results, along with Pertsova & Becker (2021), offer some support that learning mode (e.g., implicit/explicit) can affect the rules extracted and the performance on certain tasks (Moreton & Pertsova, 2023). They illustrate that ALL tasks are conducive not just to implicit learning, but also explicit learning, in ways that impinge crucially on the evaluation of their results. We echo Moreton & Pertsova (2023) on the importance of distinguishing learning strategies in the theoretical evaluation of ALL results.

At the same time, however, rule-based participants were *not* significantly more likely to correlate with *higher overgeneralization rates* ($t(45)= .15278$, $p = .88$; Figure 5). This pattern is consequential in view of the above: while explicit learning can affect learning success, it does not necessarily affect the (implicit) biases used in approaching inductive problems. An overgeneralization bias thus *inheres* in all participants, irrespective of self-reported strategy.

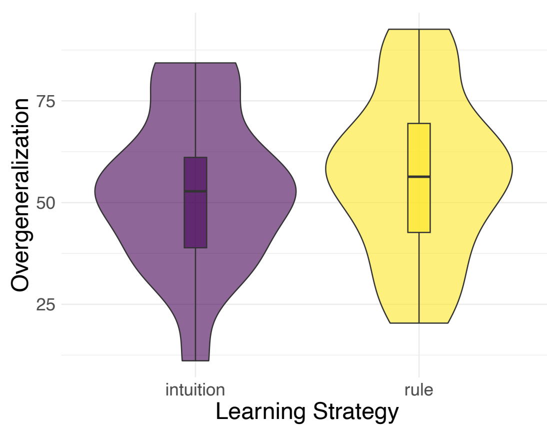


Figure 6. Overgeneralization rates by self-reported learning strategy.

3.2. *Local results* We turn now to localized, less common patterns in the dataset, which are nonetheless *consistent* with the (impoverished) input presented. We argued above that the exception-driven learning observed is plausible under explicit learning, as is participants' difficulty in adducing the (morpho)phonological generalizations at stake. This latter point is corroborated by the behavior of 9 participants, the majority from the sparser Voiced Plosives condition. These were superficially 'input unfaithful' participants, by our metric, but they nonetheless behaved in linguistically systematic ways: they consistently extracted semantic, *not* morphophonological generalizations; primarily, animacy-conditioned rules (e.g., a division between animals *vs* objects). Adults have been reported to be more likely than children to induce *semantic* over *phonological* rules, particularly when learning explicitly (Lidz & Gagliardi, 2014; Brown et al., 2021; Pertsova & Becker, 2021). Our data supports this previous work, and furthermore suggests that this semantic bias in explicit learning is strong enough to obtain despite explicit instructions in the experiment to *ignore* the semantic denotations of the visual input presented[10].

Two other idiosyncratic patterns are also worth remarking: two participants exhibited transparent instances of scope expansion, whereby diphthongization in the Voiced Plosives Condition was systematically extended to all plosives (including to held-out segments in familiarization). Finally, morphologization was also observed in one participant, who possibly morphologized the exceptional /n/ as part of the suffix (viz. their self-report "plural words end with either 'wok' or 'nok'").

This variation in generalization outcomes illustrates the range of plausible hypotheses and biases that can be observed in learning scenarios of extreme input sparsity, as in ALL. Minimally, then, an important question for future research is establishing whether the kind of generalizations observed (e.g., exception-driven learning, semantic biases, scope expansion, etc.) surface to different extents if explicit learning is minimized in the experimental set-up (see §4). In other words, it is possible, given the evidence for explicit learning observed, that this exception-driven learning strategy overshadowed other prevalent inductive biases that manifest most clearly in implicit, more abstract morphophonological generalizations.

3.3. Overall discussion Returning to the theoretical predictions in §2.4, our results are consistent with a view where learners are 'aggressive' in generalizing (RQ1). They support simplicity and scope expansion accounts, but question the empirical reality of highly conservative biases. This follows insofar overgeneralization massively predominates, with undergeneralization being largely absent. Specifically, exception-driven learning is in line with the Tolerance Principle's predictions. Our data is also in principle *compatible* with a scope expansion bias; however, it does not directly *support* it. We believe stricter evaluation criteria on these biases are needed before concluding in favor or against them: this is insofar as our data is insufficient to tease it apart from a simplicity bias; the overgeneralization observed could be characterized in terms of expansion (an increase in diphthongization contexts is observed) but crucially also simplification (e.g., simplify pluralization rule). Future work should aim to improve on this design to further tease apart the role of simplicity and scope expansion in learning. This teasing-apart, although fine-grained, is important: to the extent that simplicity and scope expansion usually overlap in their predictions, one could be masking the existence of the other. It is vital to establish whether the generalization profile of learners is best modelled by simplicity or whether scope expansion, typically hard to diagnose, actually *does* characterize (some) learning behaviors in grammar induction (see, e.g., Sayeed & Vaux, 2023, on their best-fit Bayesian learner, which incorporates both simplicity and scope expansion).

The answer to RQ2 appears to be negative: no effect of condition was observed. This answer is only preliminary, however. It is plausible that this negative result is an artefact of the experiment, given the prevalence of explicit learning techniques and exception-driven learning. As noted above, under the TP, no effect of scope is anticipated at all, only the relevant θ-threshold dictates categorical patterns. It remains an open question whether the same results would have been observed if participants had been learning (largely) implicitly or if other strategies (not the exception-driven TP) had been harnessed. Indeed, we can already argue that a small effect of rule scope is observed in the Voice Plosives condition, but not in participants' overgeneralization rates (Figure 2). Recall this condition was, by an important margin, the sparsest condition, both in terms of tokens and types, and the only condition which was fully based in token/type frequency of the two suffixes. We thus anticipate extraction of morphophonological rules to be particularly hard in this condition, given both its narrow scope and the frequency distribution of the stimuli, especially under explicit learning (Perstova & Becker, 2021). This is arguably already ratified by the fact that almost all semantically-driven generalizations in §3.2 come from participants in this condition. The potential effect of scope in the specific linguistic nature of the generalizations also suggests, as a limitation of the data analysis here, that scope may have effects in other domains not discussed here: more variables should be analyzed for effects of rule scope in the future – for example, type of generalization extracted –, above and beyond overgeneralization rates at the center of this paper. Overall, these observations insinuate that further effects of rule scope in output patterns may be uncovered if these are more meticulously probed in follow-up studies.

[10] The sparser Voiced Plosive conditions feasibly facilitated this animacy-conditioned outcome. The condition is even more impoverished in relevant than the other conditions, and, partly as a result, it is additionally unique in being the only condition balanced in *both* token and type frequency. Absent enough cues in the input to extract a morphological or phonological rule, it is plausible that adult explicit learners defaulted to more accessible linguistic variables, such as animacy (see, e.g., Pertsova & Becker, 2021, on how phonological generalizations are harder to extract than semantic ones during explicit learning).

4 Conclusion and outlook

This paper has introduced a novel experiment investigating the role of *rule scope* in ALL. Rule scope represents a variable in morphophonological rule structure and behavior that has gone understudied, relative to conservativeness or simplicity. We presented a novel programmatic experimental paradigm, based on the Poverty of the Stimulus method, that probes the effect of scope in a two-fold manner: first, we introduced three conditions that *manipulate* the scope of a rule in an artificial language, to probe its effect on (over)generalization patterns; second, we investigated the extent to which a bias for *scope* expansion is observable in the output data, privileging a maximally wide application domain of the relevant morphophonological alternation. Our results supported a view of learners as overgeneralizers in the face of extreme input sparsity. However, our data is not informative enough yet regarding a possible role of a scope bias. Elsewhere-rule preferences may have overshadowed any effect of scope; we argued this is conceivably due to the predominance of explicit learning. Nonetheless, several important implications of the data emerged: (i) no effect of condition is observed, overgeneralization of /wok/ is similarly high across conditions; (ii) specifically, we observed this is because participants most often induced rules that determined a 'default' exponent of [PL], storing perceived deviations as exceptions; we argued this overextension tracked type frequency and is modellable under Yang's (2002) Tolerance Principle, providing preliminary support for its existence in (some) adult ALL contexts; (iii) we underscored the significance of learning strategies (implicit vs explicit) and their effect in task performance, although, crucially, a bias towards overgeneralization is visible regardless across participant types. Taken together, these results are most clearly compatible with simplicity and scope expansion accounts of learners, and undermine the reality of (entirely) conservative acquirers.

It remains an open challenge to devise experimental set-ups than can tease apart the effect of rule scope in grammar induction. Some future directions include improving the design to fully distinguish between the predictions of simplicity and scope expansion, which generally overlap. The relatively simple and test-like nature of the experiment appeared conducive to explicit learning; going forward, it is important to complexify or modify the design so as to minimize the incidence of explicit learning and, therefore, to more accurately probe (implicit) learning biases. Finally, child acquirers are known to generalize more implicitly (Pertsova & Becker, 2021); similar, age-appropriate experiments with children could be illuminating. Importantly, we see these open questions not as limitations of the present study but rather as productive avenues with which to instigate greater investigation on the role of rule scope in generalization. More specifically, they have the potential to elucidate finer-grained distinctions beyond the traditional conservativeness-simplicity dichotomy. We thus finish with the note that treating rule scope as an experimental and analytical variable in morphophonological learning has significant, almost unexplored scientific potential.

References

Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham & Jo K. Evershed. 2020. Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods*, *52*, 388–407.

Albright, Adam & Bruce Hayes. 2002. Modeling English Past Tense Intuitions with Minimal Generalization. In *Proceedings of the ACL-02 Workshop on Morphological and Phonological Learning*, 58–69. Association for Computational Linguistics.

Bates, D., Machler, M., Bolker, B. and Walker, S. 2015 Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1–48.

Berwick, R.C. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

Biberauer, Theresa. 2019. Children always go beyond the input: The Maximise Minimal Means perspective. *Theoretical Linguistics*, *45*(3-4), 211–224.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge: Cambridge University Press.

Brown, Helen, Kenny Smith, Anna Samara & Elizabeth Wonnacott. 2021. Semantic cues in language learning: an artificial language study with adult and child learners. *Language, Cognition and Neuroscience*, *37*(4), 509–531.

Carr, Jon W., Kenny Smith, Jennifer Culbertson & Simon Kirby. 2020. Simplicity and informativeness in semantic category systems. *Cognition*, *202*, 104289

Chomsky, Noam. 1957. *Syntactic Structures.* Berlin: Mouton De Gruyter.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic*

Inquiry, 24, 229–345.
Cournane, Ailís. 2017. In defence of the child innovator. In Eric Mathieu & Robert Truswell (eds.), *Micro-change and Macro-change in Diachronic Syntax*, 10-24. Oxford: Oxford University Press.
Cournane, Ailís. 2019. A developmental view on incrementation in language change. *Theoretical Linguistics 45*(3-4), 127-150.
Culbertson, Jennier & Simon Kirby. 2016. Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, *6*, 1–11.
Daland, Robert, Bruces Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology*, *28*, 197–234.
Endress, Ansgar D., & Hauser, Marc D. 2011. The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 77–95.
Crain, Stepehen & Janet Dean Fodor. 1987. Sentence matching and overgeneration. *Cognition*, *26*(2), 123–169.
Gold, Mark. 1967. Language identification in the limit. *Information and Control, 10*(5), 447–474.
Hale, Mark. 2007. *Historical linguistics: Theory and method*. Oxford: Blackwell, 2007.
Halle, Morris. 1961. On the role of simplicity in linguistic descriptions. *Proceedings of Symposia in Applied Mathematics*, 12, 89–94. American Mathematical Society.
Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.
Hudson Kam, Carla L., & Elissa L. Newport. 2005. Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151–195.
Johnson, Keith. 1997. Speech perception without speaker normalization: an exemplar model. In Keith Johnson and John W. Mullennix (eds.) *Talker Variability in Speech Processing*, 145–166. San Diego: Academic Press.
Lidz, Jeffrey & Annie Gagliardi. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language, 90*(1), 58–89.
Moreton, Elliott, & Katya Pertsova. 2023. Implicit and explicit processes in phonological concept learning. *Phonology*, *40*(1–2), 101–153.
Müller, Max. 1890. *The Science of Language*. New York: Charles Scribner.
Nie, Yining. 2016. Modelling three stages of *r*-insertion in British English. Ms., New York University.
Nie, Yining, Ollie Sayeed, Erica Ross & Bert Vaux. 2019. *The evolution of phonological scope*. Ms., University of Cambridge.
Paul, Hermann. 1890. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.
Pertsova, Katya & Misha Becker. 2021. In Support of Phonological Bias in Implicit Learning. *Language Learning and Development*, *17*(2), 128–157.
Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee (Ed.) & P. Hopper, *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.
Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted. 2003. Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. *Proceedings of the West Coast Conference on Formal Linguistics*, 22, 101-114.
R Core Team. 2024. R: A language and environment for statistical computing.
Roberts, Ian. 2007/2021. *Diachronic syntax*. Oxford: Oxford University Press.
Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
Ross, Erica. 2011. *Diachronic developments in the scope of phonological developments*. Doctoral dissertation, University of Cambridge.
Sayeed, Ollie & Bert Vaux. 2023. Sound change and the evaluation metric. Paper presented at *MorrisHalle@100*, MIT.
Schuler, Kathryn. 2017. *The Acquisition of Productive Rules in Child and Adult Language Learners*. Doctoral dissertation, Georgetown University.
Stemberger, Joseph & Barbara Bernhardt. 1997. Phonological constraints and morphological development. In Elizabeth Hughes, Mary Hughes & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 602–614. Somerville, MA: Cascadilla Press.
William, Snyder. 2007. *Child language: The parametric approach*. Oxford: Oxford University Press.
Tenenbaum, Joshua. 1999. *A Bayesian framework for concept learning*. Doctoral dissertation, MIT.
Vaux, Bert. 2008. Determining the scope of generalisation in phonological acquisition: (How) are phonological generalisations acquired? NAPhC 5, 11 April 2008.
White, James. *Bias in Phonological Learning: Evidence from Saltation*. Doctoral dissertation, University of California, Los Angeles.
Wilson, Colin. 2003. Experimental investigations of phonological naturalness. *Proceedings of the West Coast Conference on Formal Linguistics*, 22, 101–114.
Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.