Simplifications made early in learning can reshape language complexity: an experimental test of the Linguistic Niche Hypothesis

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

Abstract

Languages spoken in larger populations seem to be relatively simple. One possible explanation is that this is a consequence of the simplifying influence of non-native speakers: adult learners tend to reduce complexity during learning, and large languages tend to have a higher proportion of non-native speakers. This Linguistic Niche Hypothesis, that languages adapt to their social niche, receives some statistical support from typological studies which show negative correlations between population size or number of non-native speakers and morphological complexity. Here I report an experimental test of this hypothesis, using two artificial language learning experiments to explore the impact of simplifications made by non-native-like early learners on morphological complexity. These experiments show that the presence of non-native-like early learners in a population can lead to the simplification of that language's morphology as a result of inter-generational language transmission, providing experimental support for the Linguistic Niche Hypothesis.

Keywords: Language complexity; linguistic niche hypothesis; adult learning; artificial language learning; iterated learning

Introduction

Languages differ in how they exploit particular structural devices to solve the recurring communicative problems that humans face. While this diversity presumably reflects the independent historical trajectories of different languages, there are prominent claims in the literature that at least some of this variation is a consequence of systematic differences in the structure and composition of different linguistic communities. Probably the best-known such suggestion arises from the observation that languages spoken in larger populations seem to be relatively simple, particularly in their morphology (Wray & Grace, 2007; Trudgill, 2011). One possible explanation is that this is a consequence of the simplifying influence of non-native speakers: large languages tend to have a higher proportion of non-native speakers, and adult learners tend to reduce complexity during learning; the simplicity of large languages may therefore reflect the simplifying preferences of adult learners, either because the simplifications made by those learners form part of the linguistic input to subsequent generations (Lupyan & Dale, 2010; Dale & Lupyan, 2012) or because native speakers adapt to non-native speakers during interaction (Bentz & Winter, 2013; Frank & Smith, 2018).

This hypothesis that languages adapt to their social niche, known as the Linguistic Niche Hypothesis (Dale & Lupyan, 2012), receives support from typological studies which show negative correlations between population size or number of non-native speakers and morphological complexity (e.g. Lupyan & Dale, 2010; Bentz & Winter, 2013; Sinnemäki, 2020, but see ongoing debate in Koplenig, 2019; Kauhanen, Einhaus, & Walkden, 2023; Koplenig, 2023). The Linguistic Niche Hypothesis has also been subjected to experimental tests using artificial language learning, exploring the impact of simplifications made by non-native-like learners on morphological complexity (Atkinson, Smith, & Kirby, 2018; Berdicevskis & Semenuks, 2022). A complementary body of work explores interactional mechanisms (Raviv, Meyer, & Lev-Ari, 2019; Raviv, Meyer, & Lev-Ari, 2020; Atkinson, Mills, & Smith, 2019; Frank & Smith, 2018; Atkinson et al., 2018, Experiment 3). For reasons of space I focus here on accounts dealing with the consequences of language transmission between groups of learners, rather than these interactional mechanisms, but both are presumably relevant.

In Atkinson et al. (2018) we investigated the simplifications made by individual learners and the potential for such simplifications to be adopted by a subsequent generation and therefore influence group-level language characteristics. In Atkinson et al. Experiment 1, learners trained on a morphologically complex miniature language simplify its morphology, with those simplifications being more prominent early in learning (after only 2 rounds of training); given adequate exposure (8 rounds of training), most learners accurately learn the complexities of the target language. Early learning in this experiment therefore mimics the critical feature attributed to non-native speakers in the Linguistic Niche Hypothesis, i.e. simplification of complex morphology. In Atkinson et al. Experiment 2 we then simulated a single episode of intergenerational transmission, training a second group of participants on the languages produced by participants in Experiment 1. In Experiment 2, contrary to our expectations, simplifications at the individual level did not reliably result in simplification of the population's language at the next generation: even when Experiment 2 learners learn from data featuring simplified forms produced by non-native-like early learners, they do not themselves produce languages which are simplified. We interpreted this as a consequence of mixing the output from multiple individuals, which obscures any simplification seen in the productions of individuals contributing to that mix (see also Smith et al., 2017; Kocab, Ziegler, & Snedeker, 2019). This suggests that the potential influence of non-native speakers on overall language complexity might in

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.*

fact be limited, casting doubt on whether non-native speaker simplification can be the mechanism behind the observed correlations between language complexity and population size, as envisioned in the Linguistic Niche Hypothesis.

However, in Atkinson et al. we collected a relatively small sample (48 participants in the crucial Experiment 2, across 4 conditions in a 2x2 design), meaning that the absence of significant effects should be interpreted cautiously. Furthermore, other work shows that simplifications made early in learning *can* spread. In Atkinson et al. we simulated only a single generation of language transmission, but multiple generations of transmission are known to amplify weak biases in learning (e.g. Reali & Griffiths, 2009). Berdicevskis and Semenuks (2022) therefore use a multi-generation iterated learning paradigm to show that artificial languages transmitted in chains featuring early learners simplify more rapidly than languages from chains featuring no early learners. While this result provides support for the hypothesis that simplifications by adult learners could ultimately lead to language simplification, Berdicevskis and Semenuks only run chains where each learner is exposed to the output of a single individual at the previous generation, meaning that the mixing effect identified in Atkinson et al. is not at play.

Here I use crowdsourcing to replicate Experiments 1 and 2 from Atkinson et al. with a larger sample and (in Experiment 2) exploring a wider range of input compositions. My Experiment 1 replicates the Atkinson et al. Experiment 1 result that learners trained on a morphologically complex miniature language simplify its morphology early in learning, but given adequate exposure accurately learn the target language. My Experiment 2 (N=522) shows that learners who simplify a complex morphological system (as adult learner, non-native speakers of a language might) can indeed drive language simplification through their influence on the input to subsequent generations of learners; this effect is visible even in a single generation of transmission, and even in the presence of the mixing effects identified in Atkinson et al..

Experiment 1

Experiment 1 is a conceptual replication of Atkinson et al. (2018) Experiment 1, with adjustments for online data collection, and is intended to verify that errors early in learning lead to the simplification of complex morphology.

Method

Participants 94 participants completed at least 2 rounds of the experiment. Participants were recruited via Amazon Mechanical Turk, and were native speakers of English based in the USA. Participants were paid $2-$8 ($1 per round).

The target language Participants were trained on an artificial language which provides descriptions for scenes involving animals performing movements (see Table 1). Each scene features either one or two animals performing one of three different motions (straight motion, bouncing, or looping). There are 18 possible scenes (2 numbers x 3 nouns x 3 verbs), see examples in Figure 1.

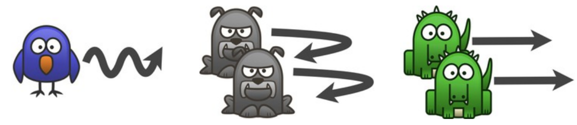

Figure 1: Three example scenes to be described in the artificial language. In Experiment 1, the corresponding descriptions in the target language would be *wona twito boingan*, *sumak grolop looponk*, and *sumuk snapop wooshesp*.

Each scene is described by three words: a quantifier giving number, a noun describing the animal, and a verb describing motion. Each word is made up of a stem and a suffix. The quantifier stems are *won* (for 1) and *sum* (for 2); the noun stems are *snap* (crocodile), *grol* (dog), and *twit* (bird); the verb stems are *woosh* (straight motion), *boing* (bouncing) and *loop* (looping). The stems were designed to be easily learned by English-speaking participants, allowing them to focus on learning of the morphological system expressed by the suffixes. The suffixes are designed to be challenging to acquire, redundantly mark number on all three lexical items, and include a mix of regular and irregular forms.

Table 1: The target language in Experiment 1. Hyphens are included here for readability, but were not included in the strings shown to participants. For quantifiers, *-a* and *-ak* mark singular and plural for dogs and birds; exceptionally, the quantifiers for crocodile mark number with *-u* and *-uk*. For nouns, *-o* and *-op* mark singular and plural, but for bird *-o* is used for both numbers. For verbs, *-an* and *-asp* mark number for dog and bird, while crocodile uses *-en* and *-esp*; an additional exception arises where the looping motion occurs with a plural image, in which case all verbs take *-onk*.

Description in the target language			English gloss
won-a	grol-o	woosh-an	“one dog straight”
sum-ak	grol-op	woosh-asp	“two dogs straight”
won-a	grol-o	boing-an	“one dog bounce”
sum-ak	grol-op	boing-asp	“two dogs bounce”
won-a	grol-o	loop-an	“one dog loop”
sum-ak	grol-op	loop-onk	“two dogs loop”
won-a	twit-o	woosh-an	“one bird straight”
sum-ak	twit-o	woosh-asp	“two birds straight”
won-a	twit-o	boing-an	“one bird bounce”
sum-ak	twit-o	boing-asp	“two birds bounce”
won-a	twit-o	loop-an	“one bird loop”
sum-ak	twit-o	loop-onk	“two birds loop”
won-u	snap-o	woosh-en	“one crocodile straight”
sum-uk	snap-op	woosh-esp	“two crocodiles straight”
won-u	snap-o	boing-en	“one crocodile bounce”
sum-uk	snap-op	boing-esp	“two crocodiles bounce”
won-u	snap-o	loop-en	“one crocodile loop”
sum-uk	snap-op	loop-onk	“two crocodiles loop”

Procedure The experiment was coded in JavaScript. Participants completed between 2 and 8 rounds; from round 2 onwards, at the end of each round participants had the option to stop and be paid, or continue to the next round.[1] On each round, 9 of the 18 possible scenes and their descriptions were randomly selected as training items. Each round consisted of 9 observation trials and 9 comprehension trials on these scenes, and 18 test trials covering the entire set of scenes.

On each observation trial, participants passively observed the scene and description (presented as text) for 6 seconds. On each comprehension trial, they were prompted with a description and asked to select the corresponding scene from an array of 4 scenes. Participants received feedback indicating whether their choice was correct or incorrect, then saw the description plus correct scene for a further 3 seconds. In the test trials, they were prompted with all 18 scenes in a random order and required to provide a description for each, by free typing into a text box, with no feedback provided.

Coding of responses Experiment 1 yielded 9682 participant-produced descriptions. Stems and suffixes were segmented automatically, followed if necessary by a manual check and re-segmentation. 2940 descriptions required a manual check, mainly produced early in learning (round 1, 1302; round 2, 733). Round 1 descriptions were excluded from analysis, and trials which could not be segmented or where the participant used the incorrect stem (e.g. *boing* instead of *woosh*) were coded as NA.

Measure of language complexity The main measure of interest is the complexity of the language participants produce in the test phase in each round of learning, specifically the complexity of conditioning of suffixes. This conditioning is complex in the target language, depending on 2 semantic features (number and animal) for quantifier and nominal suffixes, and 3 (number, animal and movement) for verbal suffixes. In a simplification of this language, conditioning might be based on fewer semantic features — e.g. a learner who has not yet learned that *twit* shows exceptional number agreement might produce a simpler regular system where nominal suffixes depends only on number. I use the complexity measure from Atkinson et al. (2018) to capture complexity of conditioning, which involves identifying the best-fitting model for each suffix type for each participant, and evaluating how many semantic features that model uses to predict suffix choice; the number of features is the complexity measure.[2]

Results

Participants improved in their accuracy in identifying the correct scene when provided with a description in comprehension trials, and in test trials increasingly accurately reproducing the target stems and suffixes.[3] This pattern of improvement over rounds matches the results reported in Atkinson et al., and indeed round 8 stem and suffix accuracy are essentially identical to the values there.

The complexity of participants' productions is shown in Figure 2. Since complexity only ranges from 0 to 4, this data was analysed using ordinal regression (using the ordinal package in R, Christensen, 2015), with round as a fixed effect and by-participant random intercepts and random slopes for round.[4] The effect of round was significant (b = 0.43, SE = 0.05, $p < .001$), indicating that participants' conditioning of suffixes becomes more complex over rounds; as can be seen in Figure 2, in round 2 participants significantly undershoot the complexity of the target language; by round 8, they reasonably closely approximate it. This confirms the finding of Atkinson et al. (2018) that in this paradigm errors made early in learning tend to result in simplification of complex morphology, a central assumption regarding the behaviour of non-native speakers in the Linguistic Niche Hypothesis.

Experiment 2

The goal of Experiment 2 is to explore whether the simplifications seen early in learning in Experiment 1 can drive language simplification in a subsequent generation of language learners, as predicted by the Linguistic Niche Hypothesis. Following Atkinson et al. (2018) Experiment 2, I simulate

[1] The experiment is challenging, particularly in early rounds, and in a pilot fixed-duration 8 round experiment many participants abandoned early. Allowing a formal mechanism for early stopping simplified running the experiment and ensured all participants were compensated fairly and straightforwardly. Of 94 participants, 17 completed only 2 rounds; 49 completed all 8 rounds.

[2] For a given participant's data for a given suffix type (quantifier, noun or verb) at a given round, I use a multinomial regression (multinom function in R, Venables, Ripley, & Venables, 2002) to predict suffix choice based on semantic features (number, animal, movement). I run all possible models featuring all combinations of semantic features and their interactions, and then select the model with the lowest AIC as the best-fitting model. I then count how many semantic features that model includes: this complexity score ranges from 0 for models which include only an intercept, e.g. if the participant only uses a single suffix, to 3 for participants whose data is best explained by a model which refers to number, animal and motion (as is the case for verbal suffixes in the target language). Atkinson et al. (2018) provide further details and examples. I made one change to the measure from Atkinson et al. (2018). The original measure treats the complexity of a quantifier suffix which is conditioned only on number (e.g. singular quantifiers end in -o, plural quantifiers end in -op) as 1 (similarly, the complexity of a nominal suffix which depends only on animal, or a verbal suffix which depends only on movement, have complexity 1). However, this is indistinguishable from a system which lacks inflection and simply has longer un-inflected stems (i.e. mono-morphemic *wono* and *sumop*). I therefore adjust the complexity scores in these cases to 0, i.e. no inflection, yielding a more conservative measure of complexity. Results using the original measure are however similar; the largest differences are at the early rounds, where this kind of simplification is most frequent.

[3] Comprehension accuracy, mean proportion correct at round 1 = 0.79 (SD=0.23); at round 8, mean = 0.97 (0.06). Stem/suffix accuracy is evaluated by normalised Levenshtein string edit distance between the target form and participant's form; error of 1 indicates no correspondence, error of 0 indicates perfect correspondence, forms coded as NA received an error score of 1. Stem error at round 2, mean = 0.41 (0.27); at round 8, mean = 0.02 (0.09). Suffix error at round 2, mean = 0.66 (0.21); at round 8, mean = 0.09 (0.14).

[4] In all models reported here I also include fixed effects for suffix (sum-coded) and all interactions involving suffix, plus appropriate by-participant slopes. Suffixes are expected to differ, since they do in the target language; however, those differences are not theoretically interesting, and I do not report effects involving suffix.

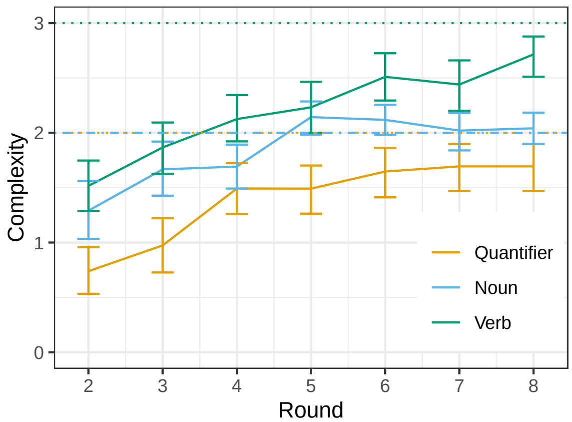

Figure 2: Experiment 1, complexity (number of conditioning semantic features) by suffix type. Lines give means, error bars give bootstrapped 95% confidence intervals. Horizontal dashed lines give complexity in the target language (2 for quantifier and nominal suffixes, 3 for verbal suffixes).

a single episode of inter-generational transmission; I take the languages produced by Experiment 1 participants (drawing data from their productions early or late in learning, i.e. at round 2 or round 8) and use those to construct new data sets which a second group of participants are trained on. I manipulate two aspects of the composition of the Experiment 2 participants' input: the number of Experiment 1 participants it is drawn from (2 or 8 participants, as a simple proxy for population size, with greater variability expected in larger populations in the real world and here; these are referred to as Small and Large populations) and the proportion of data from early vs late learners (in the No Early condition all data comes from late learners; in the Mixed condition half the data comes from late learners and half comes from non-native-like early learners; in All Early all data comes from early learners; this last condition was not run in Atkinson et al.).

Method

Participants 522 participants recruited via Amazon Mechanical Turk completed at least 2 rounds under the same criteria and payment as in Experiment 1. Participants were randomly assigned to condition, number of participants per condition ranges from 78 (in the Small-All Early condition) to 95 (in Small-Mixed).

The target language Rather than being trained on an experimenter-designed language, participants in Experiment 2 were trained on the language produced by Experiment 1 participants during testing, either their Early (round 2) or Late (round 8) productions. First, I identified a pool of Experiment 1 participants whose data was sufficiently clean to use as input data for Experiment 2: these participants had a stem error of less than 0.1 in round 2, making segmentation straightforward. All such participants were eligible to provide Early data for Experiment 2; participants who had completed 8 rounds were also eligible to provide Late data. Early and Late productions from these eligible participants were segmented (manually if required) and any stem errors (but not suffix errors) were corrected; if a given description could not be segmented it was not used as a target description for that participant for Experiment 2. This yields a cleaned set of Early and (if applicable) Late data from each eligible Experiment 1 participant. Each Experiment 2 participant requires e Early and l Late sources, depending on the condition they were assigned to (e.g. in the Small-No Early condition, e=0, l=2; in Large-Mixed, e=l=4). For each participant I randomly selected e unique Experiment 1 participants from the Early pool and l unique and distinct participants from the Late pool. This generates a large pool of descriptions for each Experiment 2 participant. For each round of training I then selected 18 descriptions from this pool, one per scene, splitting scenes across contributing Experiment 1 participants as evenly as possible. Note that since this selection is random at each round, the description for any given scene (e.g. two dogs looping) might come from different contributing Experiment 1 participants at different rounds.

Procedure The procedure was essentially the same as in Experiment 1: participants completed between 2 and 8 rounds (92 completed only 2 rounds; 222 completed all 8 rounds) where each round consisted of 9 observation trials, 9 comprehension trials, and 18 test trials. The only modification to the method was that at each round the scenes for the observation and production trials were selected such that all 18 scenes occurred once (in Experiment 1, 9 of 18 scenes were seen twice, once each in observation and comprehension); this change was intended to facilitate learning.

Coding of responses Experiment 2 generated a total of 49977 descriptions. As in Experiment 1, these were auto-segmented. There were 18623 descriptions which could not be trivially automatically segmented; the bulk of these were produced early in learning, with the vast majority of round 8 descriptions being automatically segmentable (only 603 required manual checking). Manual coding of all Experiment 2 data would be a Sisyphean task, particularly given that (1) in most cases in Experiment 1 the preliminary automatic segmentation was accurate, and (2) an additional analysis of the Experiment 1 data using uncorrected auto-segmentations produced a very similar result. Consequently, I report results on the auto-segmented Experiment 2 data; however, crucially, an analysis on the manually-coded round 8 data produces the same pattern of significant effects described below.

Results

As in Experiment 1, participants more accurately interpreted and reproduced the target descriptions as they progressed over rounds.[5] The complexity of the language participants

[5]Mean comprehension accuracy = 0.68 (SD=0.29) at round 2, rising to 0.91 (0.17) at round 8; mean stem error at round 2 = 0.46

produce increases over rounds, reflecting the same progression from simplified to more complex systems of inflection seen in Experiment 1 (significant effect of round in an ordinal regression model, b = 0.24, SE = 0.02, p $<$ .001).

The main question of interest in Experiment 2 was whether participants exposed to more data generated by early learners, or data from a more diverse set of sources (as manipulated by population size), ultimately learn a simpler language — recall that Atkinson et al. (2018) found no such effect, which we attributed to the mixing of multiple morphological systems in the input to our Experiment 2 participants.

Figure 3 (left) shows the complexity of the *input* languages participants were exposed to in Experiment 2. This confirms Atkinson et al.'s finding that mixing the output of early and late learners does indeed counteract any simplifications found in the data from a single learner, at least for the conditions from Atkinson et al. (i.e. Mixed input does not have lower complexity than No Early despite including data produced early in learning). However, the new All Early condition does show a reduction in input complexity, suggesting that there is a limit to the extent to which mixing effects can cancel simplification. This is confirmed by an ordinal regression analysis of input complexity, with fixed effects of population size, proportion of early data and their interaction; proportion of early data was coded using sliding contrasts, such that the model compares No Early to Mixed and then Mixed to All Early; population size was deviation coded. The model indicates no difference between No Early and Mixed input (b=0.09, SE=0.11, p=.381), but significantly lower complexity in All Early than Mixed input (b=-0.93, SE=0.10, p$<$ .001); there was no effect of population size on input complexity (b=0.14, SE=0.08, p=.098) and no interaction between population size and proportion of early data (lowest p=.352).

Figure 3 (right) shows the complexity of the languages Experiment 2 participants produced at round 8. In contrast to the data for input complexity, output complexity is reduced when some or all of a learner's data comes from early learners. I analysed this data using ordinal logistic regression, including fixed effects of round, population size, proportion of early data, and their interactions. Proportion of early data and population size were coded as in the analysis of input complexity; round was coded such that the model intercept is for Round 8, i.e. the estimates for the other fixed effects are for languages produced at the end of training; the model contained by-participant random intercepts and random slopes for round. As well as the expected effect of Round seen in the simpler analysis above, the model shows a significant difference between No Early and Mixed input (b=-0.63, SE=0.22, p=.004), with Experiment 2 learners exposed to Mixed input converging on simpler grammars; the difference between Mixed and All Early input is not significant (b=-0.18, SE=0.22, p=.43). There is no overall effect of population size (b=0.16, SE=0.18, p=.381), but the contrast between Mixed and All Early input does interact with population size (b=0.96, SE=0.45, p=.031). Comparing models with treatment coding of population size makes this interaction somewhat clearer: for Small populations, the contrast between No Early and Mixed input is n.s. (b=-0.38, SE=0.31, p=.224) whereas there is a significant drop in complexity from Mixed to All Early (b=-0.65, SE=0.32, p=.038); for large populations there is instead a significant reduction in complexity when comparing No Early to Mixed (b=-0.87, SE=0.31, p=.005) whereas the difference between Mixed and All Early is n.s. (b=0.27, SE=0.31, p=.378). While this rests on an interaction that is somewhat marginal (p=.031), the potential dependence of the effect of Early data on population size is intriguing, in that it suggests that simplified input can have a larger effect when it is part of more diverse input, as is the case in the Large population condition here.

(0.32), dropping to 0.09 (0.22) at round 8; mean suffix error = 0.69 (0.23) at round 2, dropping to 0.29 (0.23) at round 8. Learning accuracy is generally lower in Experiment 2, presumably because the participants' input often features inconsistencies across rounds in the suffixes used in each description of a single scene.

Discussion

Across two crowdsourced experiments I replicate Atkinson et al.'s finding that learners simplify morphological systems early in learning (making early learners a suitable experimental proxy for non-native speakers), and that linguistic data featuring a mix of complex and simplified data can itself be complex (i.e. mixing is potentially a barrier to cumulative simplification). However, an order of magnitude more data nonetheless shows that individuals exposed to data containing simplifications made by early learners themselves acquire a simpler grammar. This suggests a mechanism by which early learners can drive cumulative morphological simplification, matching the predictions of the Linguistic Niche Hypothesis. There is also some evidence that the effects of early learners are more pronounced in larger populations, where learners learn from data generated by 8 rather than 2 other participants: in my Large population condition, input from 50% early learners resulted in significant simplification, whereas in small populations I only saw significant simplification when all the data came from early learners. This could offer an explanation of why both population size and non-native speakers have been linked to simplification, beyond their correlation in real-world populations; it could be that these two factors interact in learning, with larger populations being more likely to show effects of non-native speakers. This finding would benefit from further investigation, e.g. where proportion of simplified input is varied continuously and diversity among data sources is manipulated more systematically.

These results go some way towards reconciling the apparent mismatch in the experimental literature identified in the introduction, being consistent with both Atkinson et al. (2018) and Berdicevskis and Semenuks (2022) in showing mixing effects but also cumulative simplification. This also suggests an obvious follow-up experiment: the learning paradigm from this paper could be combined with an iterated learning procedure similar to that used by Berdicevskis and

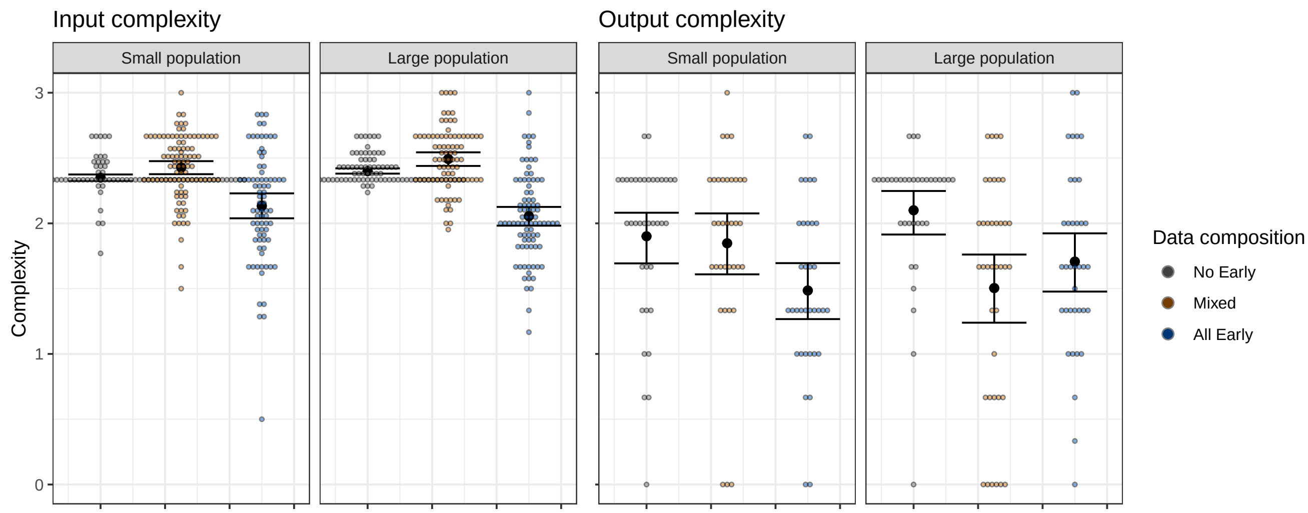

Figure 3: Experiment 2 complexity (collapsing over suffix types) by input composition. Left: complexity of input for Experiment 2 participants (averaging over rounds), showing that mixing can ameliorate simplifications made in data from individuals. Right: the complexity of round 8 languages produced in Experiment 2, showing that participants trained on data featuring some simplified input (i.e. in the Mixed and All Early conditions) produce simplified output, an effect which interacts with population size. Coloured points give by-participant means, black points and error bars give mean and bootstrapped 95% CIs.

Semenuks (2022), but running chains which feature multiple individuals per generation (as in e.g. Smith et al., 2017) to allow for mixing effects. I expect that such an iterated learning procedure would also show effects of proportion of early learner data and probably an interaction between population size and proportion of early learners, and those effects might become clearer after several iterations of transmission. Such an experiment would be resource-intensive: a 5-generation iterated learning chain with 8 participants per generation requires 40 participants; even a very modest sample of 10 chains over No Early, Mixed and All Early conditions would require 1200 participants. Computational simulation of the iterated learning process (e.g. Griffiths & Kalish, 2007; Kirby, Dowman, & Griffiths, 2007; Smith, 2009) provides another route to examining these effects, and the data reported here could provide the basis for an empirically grounded model of the inductive bases of human learners in this kind of task.

Finally it is also worth noting some suggestive connections between the role of simplification by adult learners in the Linguistic Niche Hypothesis and the broader literature. The early simulation work on iterated learning (e.g. Kirby, 2002; Kirby & Hurford, 2002), which investigated the effects of repeated episodes of language transmission on language structure, emphasised the role of language transmission through a bottleneck of limited data in driving the evolution of language structure: this bottleneck forces learners to generalise, with regularities in their generalisations eventually becoming conventions in the language. Later work (e.g. Kirby et al., 2007) reframed this as the influence of learner prior biases on the outcomes of iterated learning, with the effects of those biases being more apparent the tighter the transmission bottleneck. A reasonable interpretation of the results here is that simplifications in early learning reflect a bias for simplicity in language learning (see e.g. Culbertson & Kirby, 2016), and these results therefore fit with the general picture that a tighter bottleneck on intergenerational transmission amplifies the visibility of those biases. The predictions of the Linguistic Niche Hypothesis can therefore be seen as a specific instance of this more general phenomenon. Intriguingly, the quite distinct literature on sign language emergence emphasises the role of a different kind of early learner, children, in the creation of structure in emerging sign languages (see e.g. Senghas, 2005 for discussion, Kirby & Tamariz, 2022 for a model, and Newport, 1999, Hudson Kam, 2005 for related discussion). There may therefore be a unified explanation for both empirical phenomenon, i.e. they are a cumulative consequence of the impact of early learning. That influence might play out differently in sign language emergence and the evolution of spoken languages due to the different social dynamics in those scenarios, specifically the greater influence of child learners in the context of sign language emergence, as compared to the apparent imperviousness of established spoken languages to innovations of child learners (see e.g. Slobin, 2014). If so, the methods used here to test the predictions of the Linguistic Niche Hypothesis, combined with paradigms for studying the learning of artificial sign systems (e.g. Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019), could also provide a means to experimentally test this link between early learning and structure in sign language emergence.

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

Atkinson, M., Mills, G. J., & Smith, K. (2019). Social Group Effects on the Emergence of Communicative Conventions and Language Complexity. *Journal of Language Evolution*, *4*, 1–18.

Atkinson, M., Smith, K., & Kirby, S. (2018). Adult learning and language simplification. *Cognitive Science*, *42*, 2818–2854.

Bentz, C., & Winter, B. (2013). Languages with More Second Language Learners Tend to Lose Nominal Case. *Language Dynamics and Change*, *3*, 1–27.

Berdicevskis, A., & Semenuks, A. (2022). Imperfect language learning reduces morphological overspecification: Experimental evidence. *PLOS ONE*, *17*, e0262876.

Christensen, R. (2015). *Ordinal: Regression models for ordinal data.*

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, *6*, 1964.

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, *15*, 1150017.

Frank, S., & Smith, K. (2018). A model of linguistic accommodation leading to language simplification. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 390–395). Austin, TX: Cognitive Science Society.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*, 441–480.

Hudson Kam, C. L. (2005). Where have all the children gone? *Journal of Pidgin and Creole Languages*, *20*, 345–351.

Kauhanen, H., Einhaus, S., & Walkden, G. (2023). Language structure is influenced by the proportion of non-native speakers: A reply to Koplenig (2019). *Journal of Language Evolution*, *8*, 90–101.

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., 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., & Hurford, J. R. (2002). The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language* (pp. 121–147). London: Springer Verlag.

Kirby, S., & Tamariz, M. (2022). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*, 20200319.

Kocab, A., Ziegler, J., & Snedeker, J. (2019). It takes a village: The role of community size in linguistic regularization. *Cognitive Psychology*, *114*, 101227.

Koplenig, A. (2019). Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science*, *6*, 181274.

Koplenig, A. (2023). Still no evidence for an effect of the proportion of non-native speakers on language complexity – A response to Kauhanen, Einhaus & Walkden (2023). doi: 10.48550/arXiv.2305.00217

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLOS ONE*, *5*, e8559.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964.

Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In M. DeGraff (Ed.), *Language Creation and Language Change: Creolization, Diachrony, and Development* (pp. 161–178). Cambridge, MA: MIT Press.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20191262.

Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The Role of Social Network Structure in the Emergence of Linguistic Structure. *Cognitive Science*, *44*, e12876.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.

Senghas, A. (2005). Language Emergence: Clues from a New Bedouin Sign. *Current Biology*, *15*, R463-R465.

Sinnemäki, K. (2020). Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics*, *6*, 20191010.

Slobin, D. I. (2014). From Ontogenesis to Phylogenesis: What Can Child Language Tell Us About Language Evolution? In S. T. Parker, J. Langer, & C. Milbrath (Eds.), *Biology and Knowledge Revisited: From Neurogenesis to Psychogenesis* (pp. 255–285). Hoboken: Taylor and Francis.

Smith, K. (2009). Iterated learning in populations of Bayesian agents. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 697–702). Austin, TX: Cognitive Science Society.

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use, and the evolution of linguistic variation. *Philosophical*

Transactions of the Royal Society B, *372*, 20160051.

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford University Press.

Venables, W. N., Ripley, B. D., & Venables, W. N. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*, 543–578.