

Language Evolution in Populations: extending the Iterated Learning Model

Kenny Smith and James R. Hurford

Language Evolution and Computation Research Unit, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh,
Adam Ferguson Building, 40 George Square, Edinburgh EH8 9LL
kenny@ling.ed.ac.uk

Abstract. Models of the cultural evolution of language typically assume a very simplified population dynamic. In the most common modelling framework (the Iterated Learning Model) populations are modelled as consisting of a series of non-overlapping generations, with each generation consisting of a single agent. However, the literature on language birth and language change suggests that population dynamics play an important role in real-world linguistic evolution. We aim to develop computational models to investigate this interaction between population factors and language evolution. Here we present results of extending a well-known Iterated Learning Model to a population model which involves multiple individuals. This extension reveals problems with the model of grammar induction, but also shows that the fundamental results of Iterated Learning experiments still hold when we consider an extended population model.

1 Introduction

Language is culturally transmitted — children learn their language on the basis of the observed linguistic behaviour of others. A recent trend has been to explain the structural properties of language in terms of adaptation, by language, to pressures arising during its cultural transmission. Using this approach, properties of language such as recursion [10], generalised phrase structure [4] and compositionality [2, 8] have been shown to be adaptations which help language survive the repeated cycle of production and learning. A particular feature of this work has been its foundations in 1) linguistic theory and 2) experimentation based on computational models in the A-Life tradition.

These models suffer from an impoverished treatment of population dynamics. As discussed in Section 2.1, populations are modelled as a series of discrete, non-overlapping generations (with each generation typically consisting of a single agent), or a mono-generational collection of multiple agents. Both of these treatments of population dynamics are rather unsatisfactory, particularly given (as discussed in Section 2.2) the apparent importance of factors such as population structure and demography in language evolution in the real world. This paper has three main aims:

1. to highlight this problem and the desirability of addressing it.
2. to describe initial investigations into the impact of population dynamics on language evolution (see Section 3), highlight the problems encountered (Section 4) and present results (Section 5).
3. to outline future directions to take in addressing these issues (Section 6).

2 Population Dynamics and Linguistic Evolution

2.1 Population Dynamics in Models

Computational models of the cultural transmission of language are based around the Expression/Induction (E/I) cycle [5]. In such models, the internal knowledge of individuals determines the observable behaviour that those individuals produce (the Expression part of the process). This behaviour then forms the basis for the formation of internal knowledge in other individuals, via learning (the Induction process). This in itself is a fairly general model of cultural transmission.

Within the E/I framework, two distinct basic models have emerged which I will term the Iterated Learning Model (ILM, a term introduced by Brighton [2]) and the Negotiation Model (NM, a term based on Batali's work [1]). In both these models, agents acquire their linguistic competence based on observations of the behaviour of other agents. The difference between these two basic models is in their treatment of population dynamics.

In the NM, a population consists of a mono-generational collection of agents. At each time-step, members of the population produce linguistic behaviour, which is observed and learned from by other members of the population. There is no population turnover in the NM — new agents do not enter the population and no agents are removed.

Unlike in the NM, there is population turnover in the ILM. This turnover is typically generational, where the entire population is replaced at each time-step, although gradual turnover variants do exist. In either case, transmission is exclusively from fully enculturated individuals to naive individuals. Furthermore, the population at each generation is typically taken to consist of a single agent.¹

2.2 Population Dynamics in Language

This simplification of population dynamics is a reasonable idealisation in tackling the complex problem of modelling the cultural evolution of linguistic structure. Indeed, it would be entirely acceptable were it not for the fact that the real-world data shows that issues of population structure and demography play an important role in shaping language during language birth and language change.

Based on a review of language birth events, Sonia Ragir has suggested that “the emergence of new languages, both signed and spoken, depends on: (1) a *critical mass* of individuals generating shared patterns of linguistic practices; (2) historical continuity maintained by a continuous influx of new participants into the language pool” [13]. Similar factors have been highlighted in studies of language change driven by dialect contact. In their list of factors influencing the outcome of dialect contact, Kerswill & Williams emphasise the importance of “[t]he proportion of children to adults in the immediate post-settlement years” [7, p75] and “[t]he presence of the possibility of forming

¹ There are implementations of the ILM which move away from the single agent population model [4, 9]. However, these models either suffer from a very rigid, tightly spatially organised population model (as in [9]) or a strongly biased model of a learner (as in [4]).

new social networks among children and younger people: These possibilities are influenced by demographic factors such as high density of population, a ‘critical mass’ of population, and the presence of universal schooling” [7, p75].

3 The Basic Model

The linguistic facts outlined in the previous section highlight the importance of population dynamics in the cultural evolution of linguistic form. E/I models potentially constitute a powerful tool for the investigation of such social and population dynamics and their impact on the structure of emergent or extant languages. However, as they currently stand, the ILM and NM implementations of the E/I framework do not allow us to address such issues directly. It is our goal to develop models, building on existing models as much as possible, which are explicitly designed to allow the investigation of population dynamics and their influence on linguistic structure.

The first step in this process is to verify that the elementary results obtained through Iterated Learning experiments pertain when we move to a non-trivial model of populations — in other words, do structured languages still emerge when we move away from extremely small population models? In order to investigate this question, we have adopted the model of grammars and grammar induction developed by Simon Kirby [8, 10].² These elements of the model are briefly described below in Section 3.1. We slot this model of language and language learning into a generational turnover ILM, designed to examine the impact of learning from multiple individuals (described in Section 3.2).

3.1 Grammars and Grammar Induction

Representation An individual’s linguistic competence consists of a definite-clause grammar with attached semantic arguments. These definite-clause grammars consist of a set of rules, where the left hand side consists of a non-terminal category and the semantic label for that category and the right hand side consists of zero or more non-terminal categories, with semantic labels, and zero or more strings of characters,³ which correspond to phonetically-realised components of a signal. Semantic representations are predicate-argument structures.⁴ Two example grammars⁵ might be:

² The SICStus Prolog source code for Kirby’s model is available at <http://www.ling.ed.ac.uk/lec/software.html>

³ We consider the case where there are 26 possible characters, corresponding to the 26 characters of the Roman alphabet.

⁴ We consider the limited semantics where there are five two-place predicates and five possible arguments. As in Kirby’s models, we rule out the case where the first and second arguments are identical. This yields $5 \times 5 \times 4 = 100$ possible meanings.

⁵ Atomic semantic elements are marked with primes, characters are represented in typewriter font, upper case italics represent non-terminal categories and lower case italics represent variables over semantic elements. *S* is the privileged, top-level non-terminal category.

Grammar 1

$S / p(x,y) \rightarrow N/x V/p N/y$
 $V / \text{love}' \rightarrow \text{loves}$
 $N / \text{lynne}' \rightarrow \text{lynne}$
 $N / \text{garry}' \rightarrow \text{garry}$
plus various other lexical rules
⋮

Grammar 2:

$S / \text{love}'(\text{lynne}', \text{garry}') \rightarrow \text{lynne} \text{loves} \text{garry}$
 $S / \text{love}'(\text{lynne}', \text{beppe}') \rightarrow \text{igajojopop}$
 $S / \text{kick}'(\text{trevor}', \text{barry}') \rightarrow \text{mo}$
plus various other sentence-level rules
⋮

Both grammars would produce the string `lynne loves garry` meaning $\text{love}'(\text{lynne}', \text{garry}')$, but clearly do so in rather different ways — Grammar 1 would do so in a compositional manner (each subpart of the meaning corresponds to a subpart of the signal), whereas Grammar 2 would do so in a rote-learned, holistic manner.

Learning Learners in Kirby’s model are presented with a set of utterances, where utterances consist of meaning-signal pairs, and induce a grammar based on these utterances. Grammar induction consists of two main processes — rule incorporation and grammar compression. In an incorporation event a learner is presented with a meaning-signal pair $\langle m, s \rangle$ and forms a rule: $S/m \rightarrow s$. This amounts to simply memorising an observed utterance.

After every incorporation event, learners attempt to extract regularities and compress their grammars. This involves two main sub-processes, chunking and merging. During chunking, pairs of rules are examined in the search for meaningful chunks, which are then separated out into new syntactic categories. This leads to an *increase* in grammar size. Grammar compression is then achieved by merging similar rules. Understanding these processes is not essential for the understanding of this paper, and we refer the reader to previous papers [8, 10] for a full explanation. The key point is that learners exposed to utterances containing repeated meaningful subparts of signal will tend to arrive at grammars like Grammar 1 above, which will in turn lead to them generating utterances with regular meaning-signal correspondences. In contrast, learners exposed to an idiosyncratic set of utterances will arrive at grammars more like Grammar 2 above, and will in turn produce irregular, unprincipled sets of utterances.

Production and Reception When called upon to produce an utterance for a meaning, or to interpret a received signal, agents search, depth-first, for a combination of rewrite rules which will cover the given meaning or signal. If such a set of rules cannot be found during production then the producer applies an invention procedure, with parts of the meaning which are not expressible using the grammar being expressed with random strings. The inventor learns from its own invention, via the process described above.

3.2 The Population Model

We will begin by verifying that results obtained using this model extend to the case where populations consist of multiple individuals. We will consider a generational turnover ILM — the population consists of a set of non-overlapping generations, where each generation consists of N agents. This model of population turnover was chosen as a starting point as it most closely resembles the population model used in the majority of simulations of the emergence of linguistic structure, and can be straightforwardly extended to a more complex model involving overlapping generations.

Each individual in the population at generation n receives e exposures to the linguistic behaviour produced by agents at generation $n - 1$. A single exposure consists of a meaning-signal pair. In previous models, the set of meaning-signal pairs which an individual learns from is typically drawn from the behaviour of a single individual. In this model each learner has p cultural parents. Those cultural parents are selected at random, with replacement, from the N agents in the previous generation. Each of the e meaning-signal pairs the learner observes is then generated by a randomly selected member of this set of cultural parents.

4 Problems and Solutions

4.1 Problem: signal growth

When $p = 1$ the simulation effectively reduces to the simple case where each generation consists of a single agent. For the case where $p > 1$ (individuals potentially have more than one cultural parent), this simple extension to the generational turnover ILM runs into problems. The length of the right hand side of rules rapidly increases over generations, due to the addition of strings of meaningless terminal characters. This can result in the right hand side of rules expanding to include several hundred terminal characters. This radical string growth is a consequence of three factors:

1. the emergence of multiple, overlapping but non-identical grammars in the population.
2. inconsistent training data presented to learners (as a consequence of point 1).
3. the greedy induction algorithm.

Radical right hand side growth only occurs when you consider multiple cultural parents because the process requires variability in the input to the language learner, which does not occur in the typical single-agent ILM. The “spare” terminal characters are motivated by a small number of observed utterances. The initial addition of the spare characters is due to the greedy nature of the induction algorithm — learners do not consider multiple possible grammars at a time, nor are they capable of backtracking from the kind of over-generalisation that introduces spare characters.

4.2 Solutions to the Problem

This is clearly an undesirable artifact of the model of grammar induction, which is only exposed by moving away from the simplest possible population model. How can this

problem be resolved? The first obvious solution is to move from greedy incremental to batch grammar induction — rather than compressing the grammar as much as possible after each observation, the learner should only compress its grammar after a reasonable number of utterances have been observed. This approach has been adopted previously [15].

While batch induction may be suitable for strictly generational models of population turnover, it is unsuitable for a more realistic model of population turnover. A batch learning procedure draws a strict distinction between learners and non-learners. This kind of clear-cut distinction is undesirable in a model involving a less restrictive population dynamic — ideally, we want learners to produce observable behaviour which can be learned from by other learners. Incremental induction allows maximal flexibility in this regard.

The second possible alternative is to allow learners to entertain multiple hypotheses at any one time — rather than maintaining a single grammar, and compressing it as much as possible, a learner can maintain multiple competing grammars, compressing them differently and therefore having multiple ways of expressing certain meanings. In this scenario, we allow learners to retreat from incorrect generalisations of the type which occur in the more greedy induction process.

This multiple-competing-hypothesis approach has been used in implementations of the NM [1]. There are two main problems with this approach. The first is that there has to be pruning of grammars from time to time — maintaining all possible grammars rapidly becomes intractable. Identifying which grammars can be safely pruned is in itself a non-trivial issue. The second problem is that there must be a way of evaluating the competing grammars, in order to decide which is best. Batali implements this in his costing system. However, this costing system is rather ad-hoc. We would like the evaluation metric to be well-motivated, in terms of justifiable compression, the types of grammar evaluations that language learners make, or some other general consideration, such as the nature of the communicative task. We are working towards just such a grammar induction model, in conjunction with Simon Kirby and Willem Zuidema at the LEC.

A more immediate solution can be obtained by including justifiable production biases in the model. In other words, we allow the inducer to arrive at the wrong hypothesis, but filter out the more outrageous consequences of this hypothesis. In the case of this model, this is achievable by building in a production preference in favour of short utterances — production proceeds in a depth-first manner, as before, but at each step the producer chooses the rule with the shortest right hand side. This production bias plausibly applies in humans, as a preference for *signal simplicity* [12], or as a consequence of more general principles of least effort [14]. Including this production bias eliminates the string growth problem.

5 Results: language evolution in populations

Runs of the ILM were carried out to verify that convergence on a shared, expressive grammar was possible in the context of a population-level ILM, and to ascertain whether the number of cultural parents each individual learns from (p) has any impact on cultural

evolution. For the results presented here, we use a population size of ten ($N = 10$), and vary the number of cultural parents used in each run ($1 \leq p \leq 10$). Each learner observes $e = 50$ meaning-signal pairs, each of which is produced by one of its p cultural parents. Runs were allowed to proceed for 1000 generations. Ten runs of the ILM were carried out for each condition.

We evaluate three aspects of the populations' grammar at each generation:

Grammar Size: The number of grammar rules each agent has after learning.

Coverage: The proportion of meanings each agent can express after learning, without recourse to invention.

Communicative Accuracy: The proportion of meanings an agent is able to successfully communicate to two randomly selected communicative partners.⁶

Grammar size and coverage give an indication of the degree of structural regularity in an agent's grammar. A grammar size of approximately 50 (equal to the number of exposures each individual receives) and coverage of around 0.5 indicates an idiosyncratic, non-compositional grammar, where each meaning is associated with an unstructured signal in a holistic fashion. A grammar size of around 11 and coverage of 1 indicates a highly compressed, highly expressive compositional grammar (one sentence-level rule, plus five lexical rules for predicates and five lexical rules for arguments). Coverage of 1 also indicates stability — if all meanings can be expressed, then the underlying grammar must exhibit regularities, which subsequent learners can extract. Communicative accuracy indicates the degree of within-generation coherence in the population. A communicative accuracy of 1 indicates that agents agree on the signals which should be used to express each possible meaning.

It should first be noted that, for all values of p (numbers of cultural parents), the majority of runs converged on compressed, expressive grammars which resulted in high intra-generational communicative accuracy. In other words, fundamental results from the ILM extend to the case where the population dynamics are non-trivial — compressed, compositional, communicatively-functional grammars can evolve through Iterated Learning in populations where learners learn from multiple cultural parents.

p does appear to have some effects on the structure of the emergent grammars. Fig. 1 plots mean grammar size, coverage and communicative accuracy against p . The mean values of the three measurements were calculated by averaging over the final ten generations of each of the ten runs for each condition. As can be seen from this figure, there appears to be a "sweet spot" for p , at around $p = 4$ or 5. For these values of p , coverage is at a maximum, grammars are highly compressed and intra-generational communicative accuracy is high. Away from this value of p , coverage is (fractionally) lower, grammars are larger and communicative accuracy is lower. However, this analysis is complicated somewhat by the fact that there were non-convergent runs for $p = 9$ and 10 (as suggested by the lower average coverage for these values of p). Excluding these non-convergent runs, the overall trend is weakened, although still visible.

⁶ In order to evaluate communicative accuracy between a speaker and a hearer, the speaker produces a signal for each possible meaning, and the hearer parses that signal to arrive at a meaning. If the speaker's meaning matches the hearer's interpreted meaning, the interaction is a success.

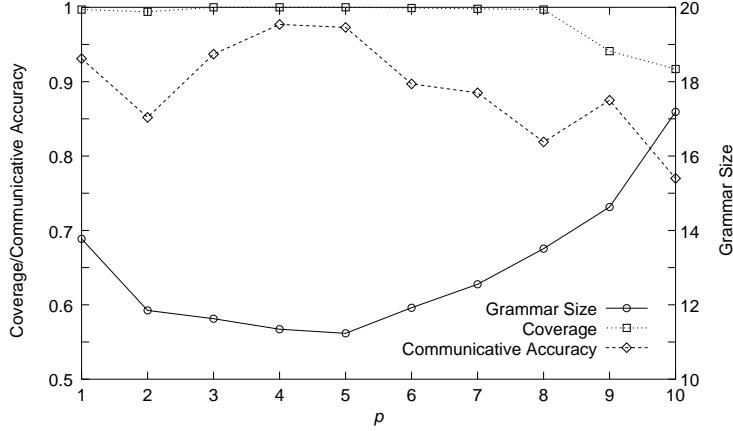


Fig. 1. The relationship between p and language structure, as evaluated by grammar size, coverage and communicative accuracy. For $p = 4$ or 5, cultural evolution in populations leads to smaller grammars and higher intra-generational communicative accuracy.

Number of cultural parents also impacts on the speed with which populations converged on a shared, stable grammar. Measuring stability in these simulations is a somewhat fraught task — the stochastic sampling of the language of the previous generation can always introduce change into an apparently stable system. The best approximation of stability is coverage — a system which can express a high proportion of meanings without recourse to invention tends to be stable. Fig. 2 plots p against time to convergence, according to two measures of stability based on coverage. As can be seen from this figure, low or high values of p tend to result in longer times to convergence, although this is perhaps less clear with respect to the stricter measure of convergence. The main point is that the values of p which were identified as the “sweet spot” in Fig. 1 tend to lead to more rapid convergence — there are certain values of p which optimise grammar compression and functionality, and lead to more rapid convergence within a population. When p is too low, learners view few alternative possible grammars and have little opportunity to preferentially acquire more compressible grammars. When p is too high, learners observe too many competing grammars, resulting in instability and difficulty in achieving consensus. At the optimal value of p , these factors are better balanced — learners witness enough variability to allow their biases to come into play, while at the same time being able to achieve stability and consensus.

6 Future Directions

We aim to expand upon this initial research in two ways. Firstly, we will develop a model of grammar induction which does not suffer from the type of problem outlined in Section 4 above. This inducer will maintain multiple potential candidate grammars, and therefore be able to retreat, to a degree, from inappropriate generalisations.

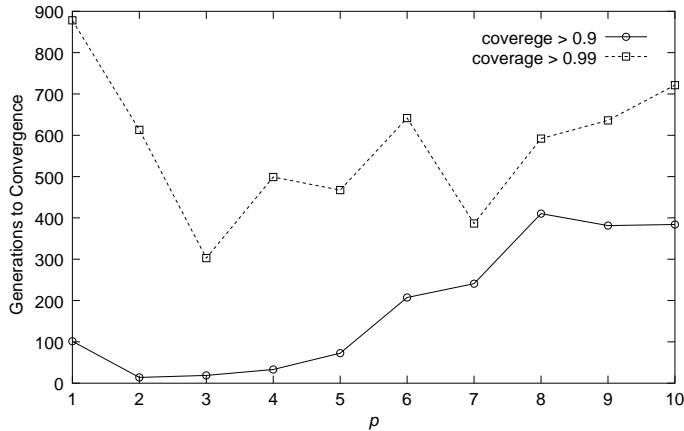


Fig. 2. Time to convergence as a function of p . A population was considered to have converged if it arrived at grammars with a certain level of coverage (either 0.9 or 0.99), and stayed above that level of coverage for the remainder of the simulation run. At least 80% of all runs reached convergence according to the weaker definition, and 70% according to the stronger definition. Once again, there appears to be a sweet spot for p , which leads to more rapid convergence.

Secondly, we will develop more sophisticated models of population dynamics. As discussed in this paper, we have verified that it is possible for a population of agents to converge on a compressed, compositional, shared grammar through Iterated Learning. We will move away from the strictly generational turnover approach, to a situation where there is gradual population replacement, and unrestricted learning interactions. We are particularly interested in the effects of high degrees of learner-learner contact, which the literature on language birth and change suggests is a key factor.

7 Conclusions

A-Life techniques can be applied to the investigation of the cultural evolution of linguistic structure. Implementations of the Expression/Induction framework (both Iterated Learning and Negotiation Models) typically suffer from an impoverished treatment of population dynamics. This is unfortunate, given that factors such as population structure and demography appear to play an important role in the processes of language birth and language change. It is our aim to develop this modelling approach to allow the impact of population dynamics on linguistic evolution to be investigated and quantified.

As a first step in this process, we must verify that previous models can be extended to a treatment of language evolution in populations, and that the key results derived from these models pertain in the new situation. The model outlined in this paper demonstrates that this is the case — general, compositional grammars can evolve culturally within a population. Furthermore, the number of cultural parents each learner has (one aspect of population dynamics) has some impact on the structure of the emergent languages and

the speed with which they evolve. However, in addition to this positive result, extension to a population model also reveals some flaws in a well-known Iterated Learning Model, which are not exposed in a minimal population model.

We are developing our approach along two lines. Firstly, we are involved in tackling the problems arising from incremental induction and greedy compression, by working on a new, theoretically well-motivated model of grammar induction. Secondly, we are extending our model to a wider range of population dynamics. This kind of experimentation, drawing on A-Life techniques, promises to be profitable in investigating language evolution in general, and, in particular, in identifying important factors in language evolution in populations.

References

1. J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Briscoe [3], pages 111–172.
2. H. Brighton and S. Kirby. The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In Kelemen and Sosík [6], pages 592–601.
3. E. Briscoe, editor. *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, 2002.
4. J. R. Hurford. Social transmission favours linguistic generalization. In Knight et al. [11], pages 324–352.
5. J. R. Hurford. Expression/induction models of language evolution: dimensions and issues. In Briscoe [3], pages 301–344.
6. J. Kelemen and P. Sosík, editors. *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life*. Springer-Verlag, Berlin, 2001.
7. P. Kerswill and A. Williams. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society*, 29:65–115, 2000.
8. S. Kirby. Syntax out of learning: the cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J. D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life: Proceedings of the 5th European Conference on Artificial Life*. Springer, Berlin, 1999.
9. S. Kirby. Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In Knight et al. [11], pages 303–323.
10. S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Briscoe [3], pages 173–203.
11. C. Knight, M. Studdert-Kennedy, and J.R. Hurford, editors. *The Evolutionary Emergence of Language: Social Functions and the Origins of Linguistic Form*. Cambridge University Press, Cambridge, 2000.
12. R. W. Langacker. Syntactic reanalysis. In C. N. Li, editor, *Mechanisms of Syntactic Change*, pages 57–139. University of Texas Press, Austin, TX, 1977.
13. S. Ragir. Constraints on communities with indigenous sign languages: clues to the dynamics of language origins. In A. Wray, editor, *The Transition to Language*, pages 272–294. Oxford University Press, Oxford, 2002.
14. G. K. Zipf. *Human behaviour and the principle of least effort : an introduction to human ecology*. Addison-Wesley, Cambridge, MA, 1949.
15. W. H. Zuidema. Emergent syntax: The unremitting value of computational modeling for understanding the origins of complex language. In Kelemen and Sosík [6], pages 641–644.