Stability conditions in the evolution of compositional languages: issues in scaling population sizes

Paul Vogt

Induction of Linguistic Knowledge group, Tilburg University, the Netherlands
Language Evolution and Computation unit, University of Edinburgh, UK.

Short title: Evolution of compositional languages in large populations

Abstract

This paper investigates the effect of scaling the population size in a simulation studying the emergence and evolution of compositionality in languages. The simulations are based on multi-agent systems that play language games in order to communicate, invent and learn language. The language games are integrated with the iterated learning model that simulates a population turnover, where the population contains adults and children. Experiments show that when the population size is increased, after an initial decrease in performance, the results show an important improvement when the population size is increased further. These results are explained by a hypothesised trade off between increasing difficulties in achieving a conventionalised system and an increased likelihood of finding structures that emerge by chance when the population size increases.[1]

Keywords: Language evolution, compositionality, language games, iterated learning model

1 Introduction

In the past decade an increasing number of models on the evolution of language have been studied, see, e.g., [1, 2] for overviews. Most of these model are based on the assumption that language is a *complex adaptive dynamical system* [3]. The idea behind this assumption is that *language* itself adapts culturally to its users as if it is a complex dynamical system, which contrasts the nativist approach in which language users adapt to use and acquire language. Central to this paradigm is that language is thought to arise through self-organisation as a result of cultural interactions among individuals and individual learning.

One important area of research has focused on the emergence and evolution of compositional languages, i.e. languages with a simple grammar in which parts of its expressions map onto parts of its meanings. The models that assume language is a complex adaptive dynamical system can – with respect to language transmission – be divided into two classes: those with a vertical transmission of language (such as the *iterated learning model* [4]) and those with a horizontal transmission (such as the *language game model* [5]). In the iterated learning model (ILM), language evolves and is transmitted from one generation to the next by allowing the adult agents to speak only to children agents who learn from the adults. In the language game model, there is typically no generational turnover, so all agents are within one generation and all agents are equally likely to speak and hear. In a recent study by Vogt [6], the two models have been combined in a simulation where there was a generational turnover as in the ILM, but where all agents could speak to each other as in the language game model. In the standard ILM, compositional structures can only emerge when the experimenter imposes a *bottleneck* on the transmission, i.e. the children only learn from a subset of the adults' languages. Vogt has shown that this bottleneck need not be imposed when children

[1]Address for correspondence: Language Evolution and Computation unit, University of Edinburgh, 40 George Square, Edinburgh EH8 9LL, UK. Email: *paulv@ling.ed.ac.uk*, phone: +44 131 6503960.

`S -> aba/`$[1_r, 0_g, 0_b, 1_s]$
`S -> A/rgb B/s`
`A -> a/`$[1_r, 0_g, 0_b]$
`A -> db/`$[0_r, 0_g, 1_b]$
`B -> ka/`$[1_s]$

Table 1: An example grammar. The grammar consists of simple rewrite rules, where non-terminal nodes (S, A and B) rewrite to a word-meaning pair (rules 1, 3–5) or to a compositional structure (rule 2). In between the square brackets are the rules' meanings, which are categorical features that relate to some feature dimension (r, g, b and s), such as the red, green and blue components of the RGB colour space and a shape feature.

are also allowed to speak (i.e. when there is a combination of horizontal and vertical language transmission), because the children implicitly face a bottleneck when producing utterances about previously unseen meanings [6].

The limitation of the studies carried out so far on the emergence of compositionality is that the populations were rather small. For instance in ILMs, most studies were carried out with only 2 agents (but see [7, 8] for population sizes up to 6 and 20 agents respectively). In Steels' language model, the studies also contained only 2 agents, and in Vogt's [6] study only 6 agents were present at each time. Although studies with more agents were carried out (e.g., in [9] there were 50 agents), in such studies each agent was restricted to learn only from a limited number of other agents.

In the recently started New Ties project (see http://www.new-ties.org), we aim at evolving language and culture in societies with populations of more than 1,000 agents [10]. This paper reports on a pilot study done with the model introduced in [7]. In particular, the study investigates the effect of increasing the population size up to 100 agents on the emerge of compositionality in simulations with the combined vertical and horizontal language transmission presented in [6].

2 The model

The model is implemented in a toolkit, called THSim [11], which simulates the Talking Heads experiment [12]. The aim of the experiment is for a population of agents to develop a shared compositional communication system to communicate about the coloured geometrical shapes that make up the population's world. As mentioned, the current model combines the iterated learning model with the language game model. As part of the learning mechanisms, some grammar induction techniques are adapted that allow agents to acquire compositional structures. Below follows a very brief description of the model, for more details, consult [6, 7].

The simulation implements a multi-agent system, where at each instance in time the population of agents contains a given number of adults and children. After the population has played a given number of language games, the adults are removed, the children become adults and new children are introduced. This process then repeats. Effectively, this population flow implements the iterated learning model. In each language game, both the speaker and hearer are selected randomly from the entire population. So, unlike in the standard versions of the ILM, where all speakers are adults and all hearers are children such that the language is transmitted vertically (see, e.g., [4, 7]), the language in this model is transmitted horizontally as in [6].

Each agent starts their lifetime without any categories or linguistic knowledge. All categories and linguistic knowledge are acquired by playing language games. The linguistic knowledge is stored by the agents in a private grammar. These grammars contain two types of rules: holistic rules (such as the first rule in Table 1) and compositional rules (such as the second rule in Table 1). In holistic rules, no part of an expression is analysed in terms of any part of its meaning. In compositional rules, parts of an expression can be analysed in terms of distinct meaning parts. (In the current model, the agents can only acquire compositional rules that take two constituents.)

In the current study, the agents play a particular variant of the language game, called the *guessing game*. A guessing game is played by two agents: a speaker and a hearer. The aim of the game is for the hearer to guess the object the speaker is referring to using a verbal expression.

In a nutshell, the game is organised as follows: Both agents look at a scene displayed on one of the monitors. This scene (or context) consists of 8 randomly sampled coloured geometrical shapes. The agents then form categories such that each object is categorised distinctively from each other object in the context. If categorisation fails, the agents can construct new categories based on the feature descriptions of the particular objects. The speaker selects one object as the target and tries to produce an expression describing this target. The production is based on the rules that are in the grammar. If no expression can be produced, a new expression is created by constructing a random string of characters, which is used to construct (a) new rule(s). The hearer then tries to interpret the expression by searching its own grammar for a composition of rules that parses the expression and that is consistent with one of the categories that relate to the current context. The hearer then points at the object it guesses is the target. If this is the correct guess, the speaker acknowledges this success. Otherwise, the speaker points at the intended target, thus providing corrective feedback.

Both speaker and hearer can find multiple ways to either produce or interpret an expression. In such cases, the agents select the way that has been most effective in the past based on the weights that are attached to the rules. These weights are adapted based on the outcome of the game. If the game is a success, the weights of used rules are increased, while the weights of competing rules that could also be selected are decreased. If the game is a failure, the weights of used rules are decreased and the hearer adopts the expression with the meaning of the intended target. While adopting the expression, the hearer first tries to see if it can be generalised by breaking up the expression and meaning in two parts, thus forming a compositional rule. If this is impossible, the hearer adopts the rule holistically.

Breaking up an expression and meaning can only be done if the hearer had previously heard one or more expression-meaning pairs that share some similarity with this new pair. For instance, consider the following simplified example. If an agent previously heard the expression-meaning pair `ab-11` and then hears the expression-meaning pair `ac-10`, it can break up these pairs forming rules such as `S -> A/x B/y`, where `A/x -> a/1`, `B/y -> b/1` and `B -> c/0`. After a hearer has broken up the expression-meaning, it will perform two post-operations to generalise the language even further and to remove any redundancies.

Summarising, the guessing game implements communicative interactions between two agents. When parts of the game fail, the agents can construct or otherwise acquire new categories or linguistic knowledge. All linguistic knowledge is mediated by weights indicating the effectiveness of used knowledge. When such weights are high, the rules have been used effectively. When agents need to select among two or more competing rules, they select one using the *winner takes all* principle. This way, effective rules will be reselected more frequently and the individual grammars will tend to become shared across the population through self-organisation in a similar way ant paths are formed.

3 Experimental results

As mentioned in the introduction, this paper investigates the scalability of the current model. To this aim, a number of simulations were done where the population size N was increased from 10 agents to 50 agents with steps of 10, and one additional set of simulations with $N = 100$. At each moment, half of the population were adult agents, and the other half were child agents. In all simulations, the model was run for 100 iterations (or generations) of 214,600 guessing games each. (This odd number was derived for earlier experiments, but since one simulation takes 1–3 weeks to process on a modern PC, no attempt was made to redo the experiments with a nicer round value.) For all different population sizes, 10 trials were carried out for statistical analysis.

At the end of each simulation, the final population was tested on the language they evolved. During this test phase, all agents in the population were presented with the same 200 situations

N	C	> 0.6	> 0.7	> 0.8	> 0.9	CA
10	0.82 ± 0.03	10	10	9	0	0.91 ± 0.03
20	0.81 ± 0.02	10	10	6	0	0.84 ± 0.04
30	0.58 ± 0.25	6	3	3	1	0.63 ± 0.13
40	0.41 ± 0.31	3	3	3	1	0.52 ± 0.21
50	0.47 ± 0.36	4	4	4	3	0.50 ± 0.26
100	0.76 ± 0.36	8	8	7	7	0.70 ± 0.29

Table 2: This table summarises the results. Column 1 gives the population size N, column 2 the gives the level of compositionality C reached at the end of the simulation, averaged over 10 trials plus their standard deviations. Columns 3–6 give the number of trials which yielded a level of compositionality higher than the value indicated in the top row. The final column presents the average level of communicative accuracy CA reached at the end of the simulations together with their standard deviations.

(a context with 8 objects, including a given target). In each situation, each agent had to produce an expression about the target, which in turn each other agent had to interpret. All learning was switched off during the test periods from which two measures were calculated: *compositionality* and *communicative accuracy*. Compositionality C measures the proportion of expressions that were produced or interpreted using a compositional rule. Communicative accuracy CA measures the rate with which agents could successfully interpret each other.

Table 2 summarises the results of this experiment. The second column provides the level of compositionality C reached at the end of the simulations, averaged over the 10 trials plus their standard deviations. As the results show, after an initial downward trend from 10 to 40 agents, the results improve substantially for larger populations. Note that for increasing population sizes, the standard deviation also increases. This is due to the fact that for larger populations, compositionality emerges at a stable level in some cases, but not in all cases. Moreover, different levels of compositionality are achieved in different trials.

In order to evaluate the levels achieved, columns 3–6 provide the frequency with which different trials of the simulations achieved a given threshold. For instance, for a population size of 30, 6 trials ended with a compositionality greater than 0.6, 3 ended with $C > 0.8$, and only 1 ended with $C > 0.9$. Strikingly, for a population size of 100, 7 out of 10 trials achieved compositionality greater than 0.98; in 1 trial $C \approx 0.8$ and in the 2 other trials $C < 0.04$. So, again these results show that although compositionality initially is affected by increasing the population size, the results improve when the population size becomes larger.

The final column of Table 2 show that communicative accuracy is pretty much correlated with the level of compositionality in the language. So, given the current model, when no compositionality emerges, accuracy in communication is very low, whereas when compositionality is high, accuracy is high as well.

It is yet unclear why the performance appears to be better for larger populations. As the simulations take a long time to run (1 – 3 weeks per trial on a single computer), it is hard to set up good experiments that can investigate what exactly is going on. So, let me speculate a little on what could be going on. First it is important to realise that – with this model – it has been shown that when children speak early in life, they face an implicit bottleneck (i.e. they may need to communicate about previously unseen objects), which puts pressure on the formation of compositionality [6]. This is true for all population sizes. However, when the population is really large (i.e. $N = 100$ agents), the agents initially invent a lot of variety in the language, because different agents invent different expressions to talk about objects. Obviously, this does not improve the ease of learning the language. However, as a result of the variation, the probability that co-varying alignments occur in the signals by chance increases, because the alphabet is of limited size and there is a tendency to create short words. In addition, the environment also has a relatively high level of co-variation in different feature dimensions, so the likelihood that co-varying structures in signals co-occur with co-varying structures in the meaning space is high as well. As the agents try to exploit these co-varying structures to form compositional structures, it is likely that such structures will arise rapidly [7]. This way, the – by chance – increased occurrence of co-varying structures soon becomes internalised in compositional structures. Furthermore, since these structures become internalised in multiple agents, they become successful in communication. Due to the positive feedback loop of the guessing games, these items are reinforced; and because the agents prefer to reuse effective items, they are used even more frequently. "Hence, the more success, the more use and the more use the more success" [3] and a conventionalised system emerges through self-organisation. This property is further enhanced, because compositional rules are more generally applicable than holistic ones, and thus tend to be used more frequently.

This way, the reason why the results are so good for large populations can – in part – be explained by the increased likelihood of finding regularities in expressions. However, this does not explain why initially the results show a downward trend when the population size increases. Apparently, the positive effect of larger populations is not yet sufficient for the somewhat smaller population sizes (i.e. $N < 50$) to achieve high results. Perhaps this is due to difficulties in arriving at shared conventions in larger populations where there is a lot of variation in the language, while there is insufficient covariation to overcome these difficulties. So, there seems to be a *trade off*: On the one hand, the larger the population, the greater the difficulty in establishing a conventionalised system. On the other hand, the larger the population, the more likely it is that regularities in the signals emerge by chance, thus giving rise to compositionality. The more compositionality in the system, the more it will be used, but – more importantly – there will occur less variation in the language, which will make compositional systems easier to learn.

Although the above may provide a suitable explanation, other explanations may prove more viable. For instance, it is possible that for some of the smaller populations the simulations have run too long and that there is somewhere an optimum in the number of language games per iteration. However, in simulations where the experiments have been run for a number of language games T per iteration proportional to $N \log N$, where N is the population size and where for $N = 100$, $T = 214,600$, the results were poorer for $N \leq 50$ than the ones presented in Table 2. (Note that the relation $T \propto N \log N$ was found to hold for studies using the language games for lexicon formation [13].) Perhaps the optimum is somewhere in between, but this is hard to assess given the computational complexity of the current model. Other explanations, e.g., based on pressures to improve the communication systems for larger populations may prove viable as well, though still do not explain the initial downward trend found. It may also turn out that, when repeating the simulations more frequently, the differences appear to be less significant, which is possible since the simulations have been repeated only 10 times and different trials can lead to different results. Further research is planned to investigate possible explanations using a less computationally expensive version of the current model.

4 Conclusions

In this paper the effect of increasing the population size in simulations regarding the emergence of compositional structures in language is investigated. The simulation is based on a model that integrates the language game model of the Talking Heads experiment [12] with the iterated learning model [4]. In the model, which is described in detail in [7], agents communicate with each other and while doing so, they construct compositional structures whenever they can, but otherwise incorporate expressions holistically. Where earlier versions of the ILM required a bottleneck on the transmission of language in order to arrive at compositional structures, the current model does not as was also shown in [6]. Moreover, where earlier versions of the ILM proved difficult to scale up in terms of population size, the current study shows that – after an initial decline – the level of compositionality can increase substantially when the population size is increased, though the emergence of a stable compositional system is not guaranteed.

From the results, it is hypothesised that there appears to be a trade off between increased difficulties at arriving at a shared communication system and an increased likelihood of finding co-varying regularities in the signal-meaning space (thus giving rise to compositionality) when the population size increases. It seems that when the population size increases sufficiently the second aspect becomes strong enough to drive the emergence of compositionality, which – once it gets off the ground – is attracted to an effective and highly compositional language through self-organisation. Hence, it may be the case that languages evolve easier in large(r) populations.

References

[1] Briscoe EJ (Ed): *Linguistic evolution through language acquisition: formal and computational models*. Cambridge: Cambridge University Press 2002.

[2] Cangelosi A, Parisi D (Eds): *Simulating the Evolution of Language*. London: Springer 2002.

[3] Steels L: **The synthetic Modeling of Language Origins**. *Evolution of Communication* 1997, **1(1)**:1–34.

[4] Kirby S, Smith K, Brighton H: **From UG to Universals: linguistic adaptation through iterated learning**. *Studies in Language* 2004, **28**(3):587–607.

[5] Steels L: **Constructivist Development of Grounded Construction Grammars**. In *Proceedings Annual Meeting of Association for Computational Linguistics*. Edited by Daelemans W 2004.

[6] Vogt P: **On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children**. *Adaptive Behavior* in press.

[7] Vogt P: **The emergence of compositional structures in perceptually grounded language games**. *Artificial Intelligence* in press.

[8] Smith K, Hurford JR: **Language Evolution in Populations: extending the Iterated Learning Model**. In *Advances in Artificial Life: Proceedings of the* 7^{th} *European Conference on Artificial Life*, Lecture Notes in Computer Science: Lecture Notes in Artifical Intelligence. Edited by Banzhaf W, Christaller T, Ziegler J, Dittrich P, Kim JT, Heidelberg: Springer-Verlag 2003:507–516.

[9] Gong T, Ke J, Minett JW, Wang WSY: **A computational framework to simulate the co-evolution of language and social structure**. In *Artificial Life IX Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*. Edited by Pollack J, Bedau M, Husbands P, Ikegami T, Watson RA, The MIT Press 2004:214–219.

[10] Vogt P, Divina F: **Language evolution in large populations of autonomous agents: issues in scaling**. In *Proceedings of AISB 2005: Socially inspired computing joint symposium* 2005.

[11] Vogt P: **THSim v3.2: The Talking Heads simulation tool**. In *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)*. Edited by Banzhaf W, Christaller T, Dittrich P, Kim JT, Ziegler J, Springer Verlag Berlin, Heidelberg 2003.

[12] Steels L, Kaplan F, McIntyre A, Van Looveren J: **Crucial factors in the origins of word-meaning**. In *The Transition to Language*. Edited by Wray A, Oxford, UK: Oxford University Press 2002.

[13] Kaplan F: **L'émergence d'un lexique dans une population d'agent autonomes**. *PhD thesis*, Laboratoire d'infromatique de Paris 6 2000.