Group size effects on the emergence of compositional structures in language*

Paul Vogt[1,2]

[1] Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
[2] Vrije Universiteit Amsterdam, The Netherlands
p.a.vogt@uvt.nl

Abstract. This paper presents computer simulations which investigate the effect that different group sizes have on the emergence of compositional structures in languages. The simulations are based on a model that integrates the language game model with the iterated learning model. The simulations show that compositional structures tend to emerge more extensively for larger groups, which has a positive effect on the time in which the languages develop and on communicative success, which may even have an optimal group size. A mathematical analysis of the time of convergence is presented that provides an approximate explanation of the results. The paper concludes that increasing group sizes among humans could not only have triggered the origins of language, but also facilitated the evolution of more complex languages.

1 Introduction

One popular hypothesis explaining the origins of language is that the group sizes in which our human ancestors lived have grown beyond a critical threshold [1]. Dunbar argues that physical grooming, which is believed to be crucial for maintaining social bonds within a group, would take up too much time required for survival oriented behaviours, such as foraging. Language, Dunbar argues, could have taken over the role of physical grooming. This paper examines the effect that group size has on the evolution of *compositional* structures in languages, i.e. structures in which parts of utterances refer to parts of their meanings and the way these parts are combined.

That group size has an effect on language development has been found in a number of studies on both human and animal communication. For instance, non-human primates have larger vocal repertoires [2] and Carolina chickadees have greater vocal complexity in information structure [3] when they live in larger groups. Humans can learn to categorise phonetic categories better when they receive input from multiple speakers than when they only learn from one, because this allows the learner to generalise better on new tokens [4]. Ragir

* This research was supported by the Netherlands Organization for Scientific Research (NWO) through a VENI grant. Many thanks to Bart de Boer and Antal van den Bosch for their invaluable comments on an earlier version of this paper.

[5] has shown that spontaneously evolved sign languages have become more structured in their grammars when used in larger communities. For instance, the sign language in Martha's Vineyard, which had a population of more than 150 signers, was well formed, as is the Nicaraguan Sign Language (100 signers in 1979 up to 500 in 1995). The sign languages of Noyha (12 signers), Grand Cayman (18 signers in 1978, earlier more) and Providence (20 signers, earlier more), however, never evolved grammar [5].

It has been shown computationally [6] and mathematically [7] that when holistic languages are transmitted iteratively from one generation to the next, they can transform into compositional ones. This is provided that 1) language learners have the ability to discover and exploit regular patterns in the utterance-meaning pairs to form compositional structures and 2) learners only observe a small part of the entire language from the previous generation. The latter *transmission bottleneck* [6, 7] is crucial, because it places a pressure on compositional structures to evolve. The reason for this pressure is that evolutionary processes tend toward stable systems [8], but *holistic* languages (i.e. languages that have no structural correspondences between parts of utterances and parts of their meanings) are not stable when transmitted through a bottleneck, whereas compositional languages are. To illustrate this, suppose that an individual wants to communicate about, say, a red triangle, which it has never communicated about before. If the language is holistic, this individual cannot use any previously learnt part of the language to produce an utterance and would have to invent a new utterance. If the language is compositional and the individual has learnt a word for red and a word for triangle from hearing utterances referring to a red square or a blue triangle, this individual can combine these words to convey the meaning of a red triangle and no new words have to be invented. Consequently, compositional languages can be transmitted more stably than holistic ones.

One limitation of the iterated learning model is that it assumes a completely *vertical transmission* of languages from one generation to the next, i.e. individuals of one generation only direct their speech to individuals of the next generation. As a consequence, this model does not allow for simulating the evolution of languages in large populations, since it would be infeasible, or would simply take too long, for the languages to converge on the entire population. This problem can be solved if the iterated learning model is combined with a model of *horizontal transmission* [9].

In horizontal transmission models [10], only one generation of individuals exists who all communicate with each other. Combining the vertical transmission model with the horizontal transmission model, yields an *isotropic* transmission model [11] that contains two generations and in which the language is transmitted in all directions (adult→child, adult→adult, child→child and child→adult). In such models, but not in vertical models, compositional languages can evolve without imposing a transmission bottleneck. Instead, individuals face a bottleneck that is an implicit and natural consequence of their development when they need to speak about meanings they have not encountered before [9]. In vertical (adult→child) models, the experimenter needs to control this transmission bottleneck to prevent children from hearing the entire language.

This paper investigates the effect of group size on the emergence of compositionality using a model based on earlier models presented in [12, 9, 11], but in which the meanings are predefined to reduce computational complexity. This model is presented in the next section. Section 3 presents the results of this study, which are discussed in Section 4. Finally, Section 5 concludes the paper.

2 The model

This isotropic transmission model implements a multi-agent system that contains $N/2$ adult agents and $N/2$ child agents, where N is the total group size. The whole group plays T rounds of language games [10], after which all adults are removed, all children become adults and new children are introduced (cf. the iterated learning model [6]). The world of the agents contains $M = 81$ compound meanings, which are constructed in 2 dimensions (e.g., colour and shape) of $m = 9$ values each (so a meaning could be something like a 'red square').

1	S → toma/[green,square]	0.2
2	S → A/colour B/shape	0.8
3	A → ba/[red]	0.6
4	B → ke/[triangle]	0.7

Fig. 1. This example grammar contains rules that rewrite a non-terminal into an utterance-meaning pair (1, 3 and 4) or into a compositional rule that combines different non-terminals (2). Whole meanings are formed by 2 features (here colour and shape). Each rule has a rule score that indicates its effectiveness in past guessing games. Only sentences of 2 constituents are allowed in this grammar.

Initially, agents' grammars are empty; the grammars are constructed by the agents playing language games (or *guessing games*). The grammar, such as illustrated in Figure 1, consists of two types of rules: *holistic rules* (rule 1) that map whole compound meanings to randomly created utterances and *compositional rules* (rule 2) that rewrite into two non-terminal rules (rules 3 and 4), each mapping meanings of one dimension to some word-form. (Note that there are two types of compositional rules in this grammar differing only in word-order.)

The grammar may contain redundant rules in that there may be different ways to encode or decode an utterance. To deal with the competition between these redundancies, each rule j is associated with a rule score ρ_j that indicates the effectiveness of the rule during past language games. When agents need to choose between a (possibly holistic) composition of redundant rules, they always select the composition i that has the highest weight w_i:

$$w_i = \begin{cases} \rho_j & \text{if holistic} \\ \rho_c \cdot \rho_{t1} \cdot \rho_{t2} & \text{if compositional.} \end{cases} \tag{1}$$

Here ρ_c is the score of a general rule, and ρ_{t1} and ρ_{t2} are the scores of the terminal rules.

In each game, two agents are arbitrarily selected from the population. One is randomly assigned the role of speaker and the other becomes the hearer. The agents are provided a shared context that contains $c = 8$ distinct meanings randomly selected from the M compound meanings. The speaker selects one meaning as the target and searches its private grammar to encode an utterance. If there are more ways to encode an utterance, the speaker selects the one that has been used most successfully in the past based on the weight of the rules. If there is no way to encode an utterance, a short random word-form is created from a finite alphabet. This new form is associated either holistically with the entire compound meaning, or with a part of the meaning if the other part is already associated with a word.

The hearer tries to decode the utterance by searching its grammar for compositions that parse the utterance such that the resulting meaning is in the context. If there are more ways to decode the utterance, the one with the highest weight is selected, yielding the meaning that the hearer guesses was intended by the speaker. If this is the correct meaning, the game is successful. Otherwise, the game fails either because the hearer guesses a wrong meaning or the hearer is unable to decode the utterances. (Note that the game's outcome is verified through explicit meaning transfer. Though this is not realistic, it is done to speed up convergence.)

Depending on the outcome of the game, the rule scores ρ_j are adapted by both agents. If the game at time t is a success, the score(s) of used rule(s) are increased by

$$\rho_j(t) = \eta \cdot \rho_j(t-1) + (1-\eta) \cdot \rho_j(t-1), \tag{2}$$

while the scores of competing rules (i.e., rules that could also encode or decode an utterance) are laterally inhibited using

$$\rho_j(t) = \eta \cdot \rho_j(t-1). \tag{3}$$

The same equation is used to inhibit the rule scores when the guessing game fails. In these equations, $\eta = 0.9$ is a learning parameter and $\rho_j(0) = 0.01$ is the initial score. These updates implement a *positive feedback loop*.

If the game fails, the speaker informs the hearer which compound meaning was intended, allowing the hearer to acquire the correct mapping. While adopting the utterance, the hearer tries to induce a compositional structure in three steps:[3]

1. The hearer searches its grammar to see if it contains a rule that can decode a part of the utterance with the correct meaning. If this is the case, the remaining part of the utterance is associated with the remaining part of the meaning. If there are more such cases, the one with highest weight is used.

[3] More details of these steps are described in [12, 9].

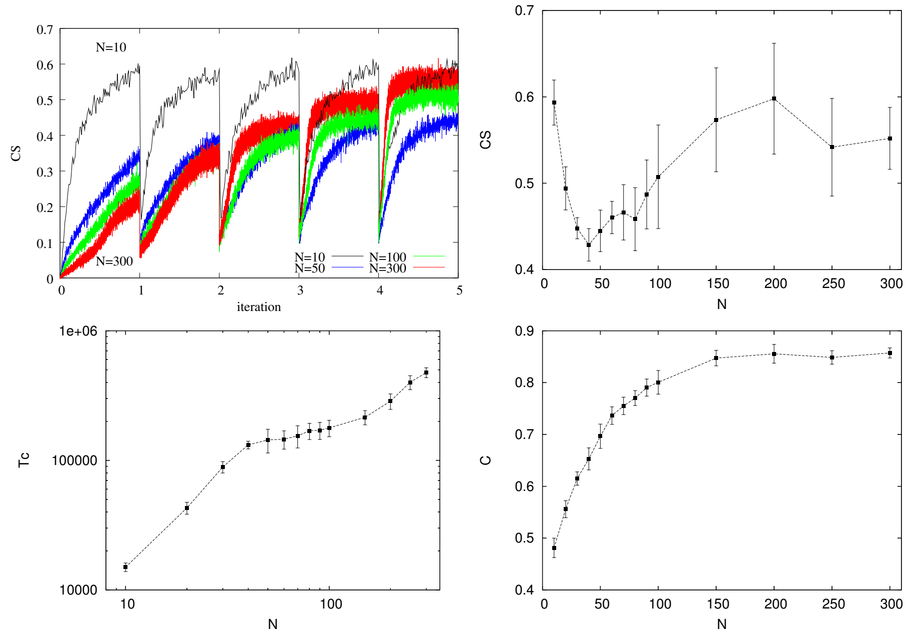

Fig. 2. Results of this study. The top graphs show communicative success CS against time in iterations (left) and against group size N (right). The bottom graphs show time of convergence T_c (left) and compositionality C (right) (both against N).

2. If this fails, the hearer searches for a *regular pattern* in the heard utterance-meaning pair compared to the most recent 1,000 instances of utterance-meaning pairs that it heard in previously played games and which are stored in a separate instance base. A regular pattern is found in two distinct utterance-meaning pairs if two utterances either start or end with the same substring and if both pairs have a similar meaning part. If such a regular pattern is discovered, the hearer breaks up the utterance and meaning following certain heuristics to form a compositional structure (see [12], pp. 221–223). The same break up is applied to all existing rules that have a similar pattern. Note that previously obtained rules are retained.
3. If the second step also fails, the utterance-meaning pair is incorporated in the grammar holistically.

Note that these induction steps are similar to those used in [6] and are inspired from usage-based approaches to human language acquisition [13].

3 Results

Figure 2 shows the results of simulating the model with various group sizes for 5 iterations of T guessing games each, where T is proportional to the group size N

approximately following $T \propto N \log N$, which was found to be the time it takes for a lexicon to converge in the population [14]. After each iteration, all adults are removed, children become adults and new children are introduced. In this study, N was varied from 10 to 300 with incremental steps of 10 between $N = 10$ and $N = 100$ and steps of 50 onwards. Each condition was repeated 10 times with different random seeds for statistical purposes.

The top graphs of Figure 2 shows *communicative success* CS, which is the fraction of successfully played guessing games during a time window of 100 games. The left graph shows CS over time (measured in iterations to scale all simulations) for a few simulations. In each iteration CS increases until its end is reached and the population is changed, through which CS drops drastically, after which it rapidly increases again to the level reached in the previous iteration and beyond, except when $N = 10$. Throughout the simulations, none of the simulations reached a value of 1, but those with larger groups show a further increase in CS. In this study, the simulations were not run longer for computational reasons[4], but it is safe to assume that after 10 or more iterations, CS would yield values near 1, as this happened in previous studies (e.g., [11]).

When setting out the average CS from the final X games against group size N (Fig. 2, top right), we see that CS first drops, then increases when $N > 50$ and finally drops again when $N > 200$. (X is roughly 10% of the number of games per iteration, T. This is done because CS can vary strongly and this average gives us more reliable values.) So, there appears to be an optimal group size around $N = 200$. Yet, although the differences between the simulations of small N and of those with larger differences (e.g., for $N = 50$, $N = 100$ and $N = 200$) are significant ($p \leq 0.01$ according to the Wilcoxon rank test), those that closely vary around $N = 200$ (i.e. $N = \{150, 250, 300\}$) are not ($p > 0.05$).

It is possible to estimate the time it takes for the curve of CS to stabilise, using a non-linear regression of its curve. This *time of convergence* T_c shows linear dependencies with N on a log-log scale, i.e. $T_c \propto N^{\beta}$, when $N < 50$ with slope $\beta = 1.58$ and when $N \leq 150$ with slope $\beta = 1.18$ (Fig. 2, bottom left). In between these values, the slope is $\beta = 0.36$. (All slopes are obtained with linear regression.) Interestingly, this result is quite different from those obtained for the evolution of lexicons, where a continuous linear dependency on the log-log scale was found. The result in the first part ($N < 50$) is similar to that of Baronchelli et al., who obtained $T_c \propto N^{1.5}$ [14, 15]. The remaining parts have slopes lower than those obtained by Kaplan, who obtained a dependency of $N \log N$ [14].

Figure 2 (bottom right) shows the evolution of *compositionality* C, which measures the proportion of two-word utterances encoded, decoded and discovered with induction step 2 (see previous section) by the population measured during the final X guessing games against group size N. It is clear that the level of compositionality increases with group size until a maximum level is reached near $C = 0.86$ for $N \geq 150$. The remainder of this paper discusses how and why group size, compositionality and time of convergence relate to each other.

[4] Processing these simulations took over 1 month using a cluster of 10 PCs

4 Discussion

The simulations reported in this paper show the surprising results that

1. compositionality tends to evolve more extensively with larger group sizes,
2. time of convergence shows different regions of dependencies with group size, and
3. there appears to be an optimum in group size concerning communicative success (though this is not a significant result).

These results are surprising, because intuitively one would think that evolving structured languages would be harder for larger populations, just as is the case for evolving (holistic) lexicons [14, 15]. However, that turns out not to be the case for all group sizes.

To explain the first two results, it is important to note that in larger groups more randomly created words occur than in smaller groups. In fact, there is a power law relation between the maximum number of randomly created words $W_{\max}$ and group size N, i.e.

$$W_{\max} \propto N^{\omega}. \tag{4}$$

The exact relation was not monitored in this experiment, but was previously observed (Vogt, unpublished) and occurs for emerging lexicons in the naming game simulations of Barronchelli et al. [15], who found that $\omega \approx 1.5$. Interestingly, they also found that time of convergence and group size had the same dependency, i.e. $T_c \propto N^{\omega}$ and argued that this was a sound correspondence. Kaplan [14], who found for a more closely related language game model that $T_c \propto N \log N$, further found that the time of convergence is linearly proportional to the number of meanings M in the language, so let us assume that

$$T_c \propto MN \log N. \tag{5}$$

Now let us assume that the number of meanings M to be acquired by the population is – on average – proportional to the level of compositionality C in the language according to

$$M = CM_c + (1 - C)M_h, \tag{6}$$

where M_c is the number of word-meaning pairs an entirely compositional language would have and M_h is the number of word-meaning pairs an entirely holistic language would have.

In this model, compositionality C depends on the chance of finding a regular pattern in both one dimension of the compound meanings and in the signals, because that is what agents find and use [12]. Since the number of meanings per dimension remains constant in all simulations, the probability of finding a regular pattern in the signals depends on the size of the alphabet Λ and the number of distinct words (or utterances) W in the language.[5] The number of words in the language follows a power law depending on group size. Hence the chance of finding and using compositional structures depends on group size.

Suppose – for simplicity – that the language in question only contains randomly constructed words that are strings containing only two letters AB, where $A, B \in \Lambda$. Let us further assume that the level of compositionality C is proportional to the probability $P(R|W)$ that, given the number of different words W that exist in a language, all possible regular patterns in word-meanings have been observed at least twice, so that for all possible compound meanings a compositional rule can be constructed.

Suppose the population has created $W = W_{\max} - 1$ different words. Now, when we find a new word starting with some arbitrary letter A, then the probability that we can find an existing word with the same letter (i.e. we can find a regular pattern R in the expression) is

$$P(R|W) = 1 - P(\neg A)^W = 1 - (1 - \frac{1}{|\Lambda|})^W, \tag{7}$$

where $P(\neg A)$ is the probability that a randomly created word does not start with the letter A. In effect, this equation says that the probability of finding a pattern is equal to one minus the probability that after creating W words none of these words start with A.

Since we assume that $C \propto P(R|W)$, we have, following Eq. (6)

$$M = P(R|W) \cdot M_c + (1 - P(R|W)) \cdot M_h. \tag{8}$$

Substituting this equation in Eq. (5) gives

$$T_c \propto (P(R|W) \cdot M_c + (1 - P(R|W)) \cdot M_h) \cdot N \log N. \tag{9}$$

[5] In the remainder of the paper, I use the term *words* for both holistic words and compositional two-word utterances, which are transmitted without a word boundary.

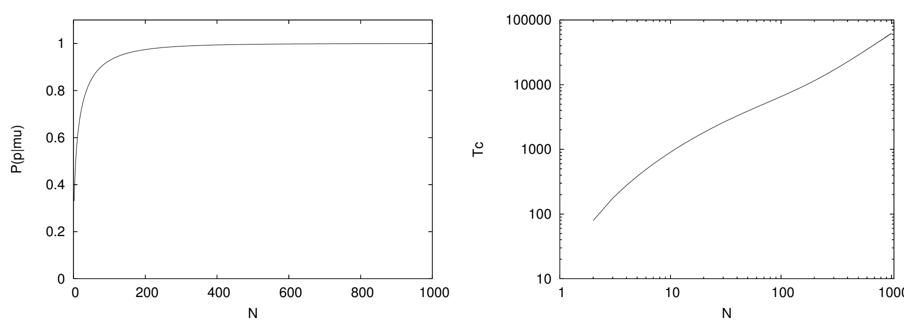

Fig. 3. $P(p|W)$ (left) and T_c (right) as a function of group size N. These figures were produced with $|\Lambda| = 4.7$, $\omega = 0.48$, $M_c = 18$ and $M_h = 81$.

Plotting Eqs. (7) and (9), using values for $|\Lambda| = 4.7$ and $\omega = 0.48$ obtained through linear regression of Fig. 2 (bottom right), produces Figure 3. This figure shows qualitatively similar dependencies for compositionality C and time of convergence T_c as observed in Figure 2. What it shows is that when the group size increases, compositionality (expressed here as $P(R|W)$) increases until it converges. While compositionality increases, the number of meanings M decrease, thus affecting time of convergence T_c by bending its curve slightly to the right and going up again, but with a lower slope on the log-log scale than previously.

The bend in the curve of T_c is less expressed than in the curve obtained with the simulation. Moreover, both $|\Lambda|$ and ω are smaller than expected, as the alphabet in the simulations contains 15 characters and – if Baronchelli et al. are right – the exponent of Eq. (4) was expected to be closer to 1, or even exceeding 1. Unfortunately, in these particular simulations, the evolution of the maximum number of words in the language was not monitored, so we have no data to measure ω based on Eq. (4).

A probable reason that the values are lower than expected is that the results on C and T_c were presented concerning the fifth iteration, before which the language had already been developed to some extent. This would increase the likelihood of finding a regular pattern, thus lowering $|\Lambda|$. As a consequence, this also decreases the need for creating new random words, thus lowering ω.

Currently, the simulations are being repeated with a faster implementation of the model than used to generate the data presented here, thus allowing to do more runs of the simulations for better statistics, to run them for a longer time and to go beyond the 300 agents. In addition, these new runs do monitor the evolution of words, so we can have a better estimate of parameters for the mathematical model. Moreover, the additional simulations aim to investigate the effect that group size has on the level of communicative success, which revealed an apparent optimum. However, this optimum was not significant. Better statistical results and larger group sizes should shed more light on this issue.

5 Conclusions

The simulations in this paper show that for larger groups compositional languages evolve more extensively due to the increased number of words, which increases the likelihood of finding regular patterns in utterance and meaning. As a result, individuals tend to use the compositional language bits more frequently, so there are less meanings to be distributed among the population, which affects time of convergence.

The relation between time of convergence and group size has three phases. First, time of convergence is increasing relatively fast with increasing group sizes, then it increases much slower, after which it starts to increase faster again, but with a slower rate than for the smaller group sizes. As shown mathematically, the first decrease in the slope coincides with a strong decrease in the number of meanings to be distributed, due to the increased compositionality. The later increase in the slope coincides with compositionality reaching a stable maximum.

So, Dunbar's hypothesis that language has originated to facilitate a different mode of grooming when the groups in which hominids started to live exceeded a certain threshold [1], whether right or wrong, seems to have an interesting consequence. Our ancestral evolution to live in larger groups did not only put a pressure on language to originate, but actually facilitated the emergence of compositional languages.

References

1. Dunbar, R.: Grooming, Gossip, and the Evolution of Language. Harvard University Press (1998)
2. McComb, K., Semple, S.: Coevolution of vocal communication and sociality in primates. Biology Letters **1**(4) (2005) 381–385
3. Freeberg, T.: Social complexity can drive vocal complexity: Group size influences vocal information in carolina chickadees. Psychological Science **17** (2006) 557
4. Lively, S., Logan, J., Pisoni, D.: Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. The Journal of the Acoustical Society of America **94** (1993) 1242
5. Ragir, S.: Constraints on communities with indigenous sign languages: Clues to the dynamics of language genesis. In Wray, A., ed.: The Transition to Language. Oxford University Press, Oxford (2002)
6. Kirby, S.: Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation **5(2)** (2001) 102–110
7. Kirby, S., Dowman, M., Griffiths, T.: Innateness and culture in the evolution of language. Proceedings of the National Academy of Sciences **104(12)** (2007) 5241–5245
8. Cavalli-Sforza, L.L., Feldman, M.W.: Cultural Transmission and Evolution: A quantitative approach. Princeton University Press, Princeton, NJ (1981)
9. Vogt, P.: On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. Adaptive Behavior **13(4)** (2005) 325–346
10. Steels, L.: The synthetic modeling of language origins. Evolution of Communication **1(1)** (1997) 1–34
11. Vogt, P.: Cumulative cultural evolution: Can we ever learn more? In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J., Marocco, D., Meyer, J., Parisi, D., eds.: From animals to animats 9: Proceedings of the Ninth International Conference on Simulation of Adaptive Behaviour, Berlin, Springer Verlag (2006)
12. Vogt, P.: The emergence of compositional structures in perceptually grounded language games. Artificial Intelligence **167(1–2)** (2005) 206–242
13. Tomasello, M.: Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press (2003)
14. Kaplan, F.: Simple models of distributed co-ordination. Connection Science **17**(3-4) (2005) 249–270
15. Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., Steels, L.: Sharp transition towards shared lexicon in multi-agent systems. Journal of Statistical Mechanics **P06014** (2006)