INTRODUCING PRESSURE FOR EXPRESSIVITY INTO LANGUAGE EVOLUTION EXPERIMENTS

ALEKSANDRS BERDICEVSKIS

Department of Foreign Languages, University of Bergen, Postboks 7805
Bergen, 5020, Norway

I present a language evolution experiment that integrates two evolutionary pressures (for learnability and for expressivity) and two directions of language transmission (vertical and horizontal). Members of each generation of a diffusion chain are given a communication task. The experiment shows that the compositionality of their languages increases over generations, but the reproduction fidelity does not. Moreover, the compositionality of the input language is a good predictor for the compositionality of the output language, but not for the reproduction fidelity. These results, together with the analysis of intragenerational alignment, allow to estimate the contribution of each kind of pressure and each transmission way.

1. Introduction

In a well-known laboratory experiment on language evolution by Kirby et al. (2008) structure emerges in initially unstructured languages, as they are transmitted over generations of a diffusion chain, obviously adapting to the pressure for learnability. The same experiment, however, shows that this pressure alone is not enough.

Only when homonyms were removed by experimenters from every input language, did the languages develop a compositional structure (otherwise they became strongly underspecified). This removal of ambiguity acted thus as a pressure for expressivity, preventing the languages from losing their expressive power. While efficient, this model is artificial. My primary aim is to test whether the pressure for expressivity can be implemented in a more natural way, by using a communication task. This can increase the role of intentional design, but it is likely that the task to adhere to the learned language will temper it.

Introducing communication into iterated learning experiment also answers another call of the developing field (Garrod et al. 2010; Theisen-White et al. 2011): to take into account both horizontal and vertical transmissions and to estimate their roles in the evolution of language.

2. Methods

2.1. *Model of vertical transmission*

48 secondary school students participated in the experiment, all native speakers of Russian, mean age 15.4 years. Participants were rewarded with candies for successful rounds (see below).

Participants were asked to learn an “alien” language and to communicate in it in pairs. During the learning stage the alien language was presented to the participants on a computer screen as a set of string–picture pairs. The members of a pair were trained on the same language: the first pair in a chain on the randomly generated one, and the following pairs on the output of the preceding pair (see below). The training took place simultaneously, but in different rooms.

2.2. *Language*

The meaning space was determined by three parameters: shape (four possible values); colour (two); background (three). The design was made asymmetrical (4x3x2) in order to keep the meaning space less predictable[1]. The strings for the input language of the first pair were generated randomly in lowercase Cyrillic letters (given here in Latin transliteration).

2.3. *Model of horizontal transmission*

After the learning stage, the communication stage took place. Two participants were seated in the same room at different sides of a screen, so that they could not see each other. The only thing they knew was that their partner was an “earthian” who had received the same training. Voice communication was forbidden.

One of the participants (A) was shown a random picture on the screen of a laptop. She wrote a name for it on a paper card. The card was passed to the participant B who depicted on it a picture that, in her opinion, corresponded to that name. If the depiction coincided with the original stimulus, this round was considered successful and the participants were awarded a point. The original stimulus was then shown to B, and the card with the depiction was shown to A, so that they both knew their result. When writing a name for a figure, participants used a black pen, when drawing a picture, they could use any of four coloured pencils (red, blue, green, yellow).

For the next round, the roles were exchanged, and the next picture was

[1] The meaning space was indeed poorly predictable: participants often used green and yellow pencils (see below); at debriefings, very few were able to answer how many stimuli they saw.

shown. Such rounds were held for all the 24 stimuli (presented in random order, each stimulus once). Thus, each of the participants performed 12 operations of naming a picture and 12 operations of drawing a picture for a given name.

The communication system shared by A and B was considered to consist of the signs used in successful rounds. In other words, if the participants won a round, the name on the card was considered a word of their shared language, its meaning being the picture on the same card. This output language (E-language, as opposed to each participant's I-language) served as input for the next pair.

2.4. *Bottlenecks and failures*

If the number of signs in an E-language (*n*) was larger than 12, then *n-12* signs (chosen randomly) were removed. This rule was applied to the first input language as well. If *n* did not exceed 12, no bottleneck was applied. If *n* was smaller than 2, then language was considered extinct (evolutionary dead-end), since the pilot experiments showed that in this case the participants were unable to perform the communication task. Thus, the number of signs in the input languages was not fixed and could range from 2 to 12.

2.5. *Data analysis*

I analyze changes in the compositionality of E-languages and the fidelity of their reproduction. The expectation is that the cumulative evolution should lead to the increase of both parameters over generations.

The fidelity of reproduction shows how many signs of the E-language of the previous generation were correctly reproduced. It is calculated as transmission error substracted from one. The transmission error was measured as a mean normalized Levenshtein distance between the signals in E-language of generation *n* and the corresponding signals in E-language of generation *n-1*. If in the language of generation *n* a signal was present, but in the language of generation *n-1* it was missing, this was not considered to be an error (the signal had not been part of the input language, there was no name to reproduce), and distance was equated with 0. If in the language of generation *n-1* a signal was present, but in the language of generation *n* it was missing, distance equaled 1 (the name that had existed was lost).

Compositionality was measured using two different methods. One was the *RegMap* metric as in Tamariz et al. (2010) (originally suggested by Tamariz&Smith 2008). *RegMap* allows to quantify how regular the mappings between signals and meanings are. Partial *RegMaps* estimate how confident a learner can be that a meaning compone,nt (e.g. shape) is associated with a signal component (e.g. first morpheme). Full *RegMap* tests whether these associations are biunique, returning one value for the whole language. This value ranges from 1 (there is a biunique association for every component of both spaces) to 0 (there is no trace of such associations)[2].

Another measure was that of Kirby et al. (2008: 10686), the Mantel test. Here, the structure is measured as the Pearson's product-moment correlation between the matrix of the distances between all pairs of strings in the language (Levenshtein distance) and the matrix of the distances between all pairs of meanings[3] in the language. Significance is measured by Monte Carlo method (1000 randomizations or as many as possible for smaller languages). The variable number of signs in the languages, however, constitutes a problem: smaller languages often have high correlation coefficient, but low significance. If p-value was larger than 0.05, the matrices were considered not correlated and the measure of structure thus equal to 0. This makes the measure blind to changes in structure in these languages, so I consider RegMap as the basic measure.

Expressivity is also reported, measured as a number of signs in an E-language (the score obtained by the pair in the game).

3. Results

3.1. *Vertical transmission*

Three chains (2, 4 and 5) ended up at the first round, since the pairs produced E-language consisting of one sign only. This shows that the tasks offered to the participants were really difficult. Since the results of these pairs are not indicative of the cumulative evolution of language, they are not included into further analysis, which is thus limited to the "successful" chains.

Three pairs struggled successfully through the first round, which resulted into three chains (1, 3 and 6) of seven generations each (Fig. 1). In these chains, there is an increase of compositionality over generations within a chain, as confirmed by Page's trend test (m=3, n=8; using RegMap: L=559, p<0.01; using Mantel test: L=541, p<0.05). There is, however, no significant increase of reproduction fidelity (L=365, m=3, n=7, p≈0.44).

If we look at the dependence of different parameters of an output

---

2 *RegMap* in its current form relies on the assumption that the language has no "empty cells", that all possible meanings are expressed. This is not the case with my data, and smaller languages receive lower score, even if the signs they include are perfectly regular. I do not adjust for that, since the inability of language to express certain meanings (often present in morphosyntax of natural languages) might also be viewed as irregularity that affects the learner's perspective.

3 The meaning distance is the number of differences between the two stimuli (0, 1, 2 or 3).

E-language on the compositionality (measured using RegMap) of the input E-language, the picture is as follows. Full *RegMap* of the output language correlates strongly with the full *RegMap* of the input language of the same generation (Pearson's r=0.61, n=19, p<0.01).

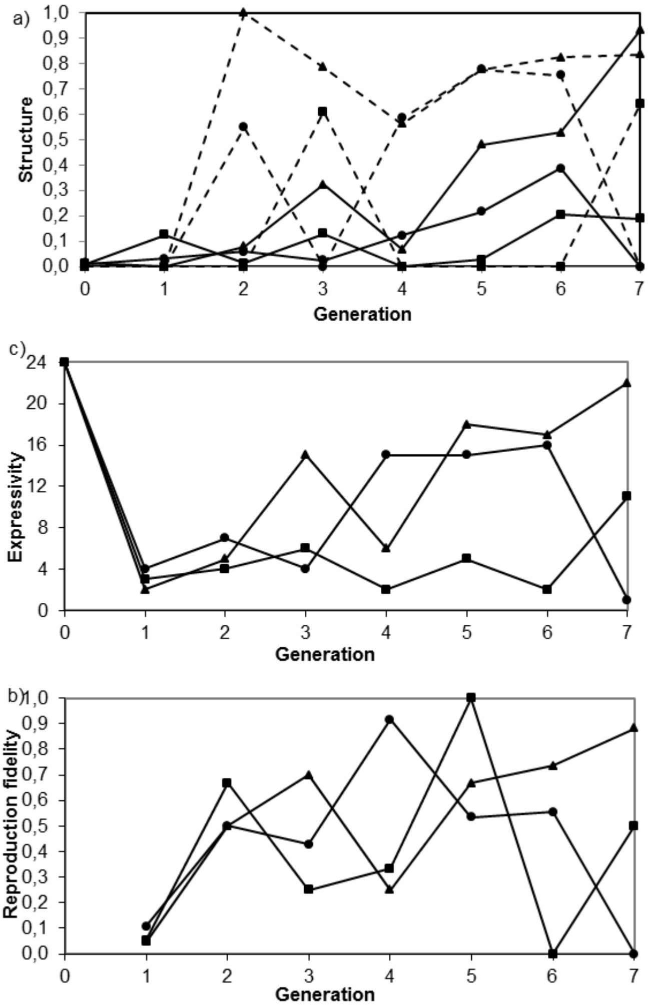

Figure 1. Parameters by generation: a) structure (solid line – RegMap, dashed line – Mantel test); b) reproduction fidelity; c) expressivity. Circles – chain 1, squares – chain 3, triangles – chain 6.

Partial *RegMaps* of the output language also correlate with the corresponding partial *RegMaps* of the input language quite strongly (r=0.53, n=175, p<0.001), thus showing that associations between meaning and signal components are rather stable. The performance of a pair (the expressivity of their E-language) is also correlated with the compositionality of their input language (r=0.45, n=19, p<0.05). The reproduction fidelity of a language, however, does not correlate with its compositionality (r=0.18, n=19, p=0.45).

Mantel test gives a similar picture: performance is correlated with structure (r=0.50, n=19, p<0.05), while fidelity is not (r=0.19, n=19, p=0.42). The only difference is that correlation between the structure of the output language and the structure of the input language is not significant (r=0.36, n=19, p=0.11).

3.2. *Horizontal transmission*

It would be reasonable to expect that within a pair an alignment of the I-languages of the individuals takes place, and they gradually converge on the same language. In other words, participants should agree upon the stimuli appearing towards the end of interaction stage better.

To test this, I calculate the meaning distances between the figure shown to one participant and the one drawn by the second participant for all the moves. If the hypothesis is correct, the distances should decrease towards the end of the chain. Linear mixed model fit confirms this is the case: t=-2.956, p<0.01, degrees of freedom (502) determined as recommended by Baayen (2008:270).

3.3. *Linguistic phenomena*

It is also interesting to look at various linguistic phenomena and types of structure that emerge. The output of the pair 6:7 (Fig. 2a) is an unambiguous agglutinative language: the first letter denotes shape, the second denotes colour. The rest denotes background in a way which may be called hypercompositional: *-hovo* is a combination of *-ho* and *-vo*, and its meaning ('filled') may also be viewed as a superposition of the meanings 'striped' and 'polka dot'. The diachrony confirms this: first, *-ho* and *–vo* became associated with resp. 'striped' and 'polka dot'. Later, 'filled' became associated with *-hovo*, while the association of *–vo* and 'striped' was lost. It resurrected after three generations of variation (*-bi*, *-vi*, *-hi*), obviously under the influence of *-ho* and *-hovo*.

The E-language of the pair 1:4 (Fig. 2b) is also agglutinative. Each syllable denotes a meaning component, 'filled' is expressed by zero morpheme. Irregularity can be observed: for example, the order of morphemes is inverted in the names for all but one filled figures: colour comes first, shape comes second.

The output language of pair 6:7 (Fig. 2c) is more unusual. It resembles the communication systems of monkeys as described e.g. by Ouattara et al. (2009): a limited number of unchangeable signs are combined in long sequences.

a)		**square**	**triangle**	**circle**	**rhomb**	b)	**square**	**triangle**	**circle**	**rhomb**
filled	**red**		tohovo	hohovo	bohovo		kapa	vuka	kale	kape
	blue	vihovo	tihovo	hihovo	bihovo		lepa	levu	lele	lepe
striped	**red**	vovo	tovo	hovo	bovo					
	blue	vivo	tivo	hivo	bivo			vunape	lenapa	penape
polka dot	**red**	voho	toho	hoho	boho			vukale		
	blue		tiho	hiho	biho		palele		lelele	pelele

c)		**square**	**triangle**	**circle**	**rhomb**
filled	**red**				no-no: l-l-l-l-l
	blue	l-no-no: l-l-l-no-no	l-no: l-l-l-no-no		no-no: l-l-l-no-no
striped	**red**				
	blue	l-no-no: l-no-no	l-no: l-no-no	l-l: l-no-no	
polka dot	**red**	l-no-no: no-no-no-no-no	l-no: no-no-nu-nu-nu	l-l: no-no-no-no-no	
	blue	l-no-no: l-no			

Figure 2. E-languages: a) chain 6, generation 7; b) chain 1, generation 4; c) chain 3, generation 7. The words are shortened: *l* stands for *levu*, *no* for *nepopo*, *nu* for *nepopu*.

4. Discussion

My results are in agreement with those of Kirby et al. (2008) with one salient exception: in my experiment, there is no significant increase of reproduction fidelity, given the increase of compositionality.

One explanation for this is that the analyzed seven generations represent the early stages of language evolution, when languages have to change in order to become more structured. It is possible that should chains 3 and 6 be continued, the reproduction fidelity will grow.

Alternative explanation can be sought in the interplay of horizontal and vertical transmissions. While vertical transmission is affected by the learning bottleneck, horizontal transmission is affected by the gap between the participants' I-languages. In order to communicate effectively, they have to bridge this gap by converging on a compositional language (and they gradually do that, as shown in Section 3.2.). The more compositional their input is, the smaller the gap is, the faster the alignment is achieved. In this case, it is possible that the participants succeed in reaching compositionality and expressivity, but still reproduce the input language poorly. It is less likely that they fail to reach compositionality, get a low score, but reproduce the input language well.

This is understandable: in this model, as well as in the real world, speakers' *main* goal is successful communication, not the reproduction of what they learned. People are remarkably effective in quickly establishing communication even in the unfavourable conditions, as shown both by experiments (Scott-Phillips et al., 2009; de Ruiter et al., 2009) and observation of emerging new languages (Sandler et al., 2005, Senghas et al. 2005). It might be that while compositional input makes this process more efficient, faithful reproduction of this input is less necessary. After all, languages do change.

Acknowledgements

I am extremely grateful to Eugenia Kuznetsova, Esther Sheynkman, Fedor Vinokurov and Liudmila Zaydelman for their help in conducting the experiments, and to Dmitry Logvinov for creating the necessary software.

References

Baayen, R.H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge University Press.

De Ruiter, J.P., Noordzij, M.L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S.C., Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, *11(1)*, 51–77.

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies*, *11(1)*, 33–50.

Kirby, S., Smith, K., and Cornish, H. (2008) Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceeding of the National Academy of Sciences*, *105(31)*, 10681–10686.

Ouattara, K., Lemasson, A. Zuberbьhler, A. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceeding of the National Academy of Sciences*, *106 (51)*, 22026–22031.

Sandler, W., Meir, I., Padden, C. and Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceeding of the National Academy of Sciences*, *102(7)*, 2661–2665.

Scott-Phillips, T.C., Kirby, S. and Ritchie, G.R.S. (2009). Signalling signalhood and the emergence of communication. *Cognition 113(2)*, 226–233

Senghas, R.J, Senghas, A. and Pyers, J.E. (2005). The emergence of Nicaraguan Sign Language: questions of development, acquisition and evolution. In J. Langer, S.T. Parker and C. Milbrath (Eds.), *Biology and Knowledge revisited: From neurogenesis to psychogenesis*. Lawrence Erlbaum Associates.

Tamariz, M., Brown, J.E. and Murray, K.M. (2010). The role of practice and literacy in the evolution of linguistic structure. In A.D.M. Smith, M. Schouwstra, B. de Boer and K. Smith (Eds), *Proceedings of the 8th International Conference on the Evolution of Language*, 313–320. World Scientific.

Tamariz, M., and Smith, A.D.M. (2008). Regularity in Mappings Between Signals and Meanings. In A.D.M. Smith, K. Smith and R. Ferrer-i-Cancho (Eds), *Proceedings of the 7th International Conference on the Evolution of Language*, 315–322. World Scientific.

Theisen-White, C., Kirby, S. and Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. *Proceedings of Cognitive Science Society Conference*, 956–961.