Lexical Language Evolution in Networked Human Groups

David Reitter
Department of Psychology, Carnegie Mellon University
and: College of Information Sciences and Technology, Penn State University
reitter@psu.edu

ABSTRACT
The development and refinement of natural-language communication systems among networked individuals is not well understood and difficult to study. This paper uses a task providing a controlled environment for the goal-oriented, collaborative exchange of short, natural-language messages between experimental participants (20 per group) in order to demonstrate lexical convergence. A technique for illustrating convergence based on graph layout by multidimensional scaling is described. While reliable convergence is shown, it is limited to the collaborative or communicative situation: participants did not adopt group terms in a separately administered post-test questionnaire.

Author Keywords
memory and knowledge; language evolution; natural-language dialogue; vocabulary

INTRODUCTION
The emergence of communication systems is often characterized as a system consisting of two interacting processes. The first is the genetic selection of a cognitive system providing the computational means to process language. The second involves the cultural transmission of conventions that associate meaning with symbols. The first process is difficult to observe, although cognitive architectures (e.g., [1]) and extensive work in linguistics provide computational models of the its current state. The second, horizontal process has become more observable as recent years have brought advances in methods that enable controlled experimentation [6].

In light of much work in network science, the structural properties that enable language evolution are of particular interest. Common self-organizing real-world networks often evolve into graphs with small-world properties that implement high clustering yet short average distances. Whether structural properties of the network facilitate evolutionary processes in conjunction with the specific properties of human memory is question I seek to answer in the long term.

Words and Networks: Language Use in Socio-Technical Networks. Workshop at WebSci 2012, June 22, 2012, Evanston, Illinois, USA.

Recent work [4] contrasts fully connected networks with a set of disjoint pairs or connected communities of communicators (n=8 per group), who use drawings to convey meaning (Pictionary game). Fay et al. find qualitative differences in the convergence of meaning-drawing associations adopted by participants, as well as quantitative differences in terms of task success. Does the convergence of such communication systems extend to natural language lexica? How does it interact with larger networks, which are not fully connected, or where small-world properties influence the outcome? Which label-meaning associations "win" the evolutionary game is also a pertinent question. Some models [10] predict that among several alternative labels for a given meaning, the more specific one *blocks* the use of the more general one. This prediction inspires some of the empirical investigation presented here. Here, my goal is to derive quantiative, closed-form models predicting success or failure of labels.

THE GEO GAME TASK
The Geo Game is an interactive task that involves a network of participants communicating with each other and thereby facilitating their individual and joint success in the game. This design provides an experimental model of human communities (Figure 1), where information may spread from peer to peer by word-of-mouth, and an equivalent agent-based simulation. The game is intended as a model of real-world cooperative foraging tasks. Communication paths between players are defined by the edges of a social network graph; each participant may broadcast to their network neighbors by typing short messages displayed instantly to all of them.

Players are shown a map of cities connected with roads. Using this map, they are asked to travel between the cities in order to find a specific item that is hidden in one of the cities: only when a player has arrived at a city are the locally available items revealed. The task requires them to scout the area; they are, however, most efficient if they have information about where the item is located. (Once they have found an item, a new one is assigned.) Thus, communication and individual memory are helpful. The Geo Game task forces a realistic tradeoff between directing attention to communication and to exploring the world: both activities lead to information gain, but are associated with attentional costs. This defines economic constraints for the language developed by the participant groups.

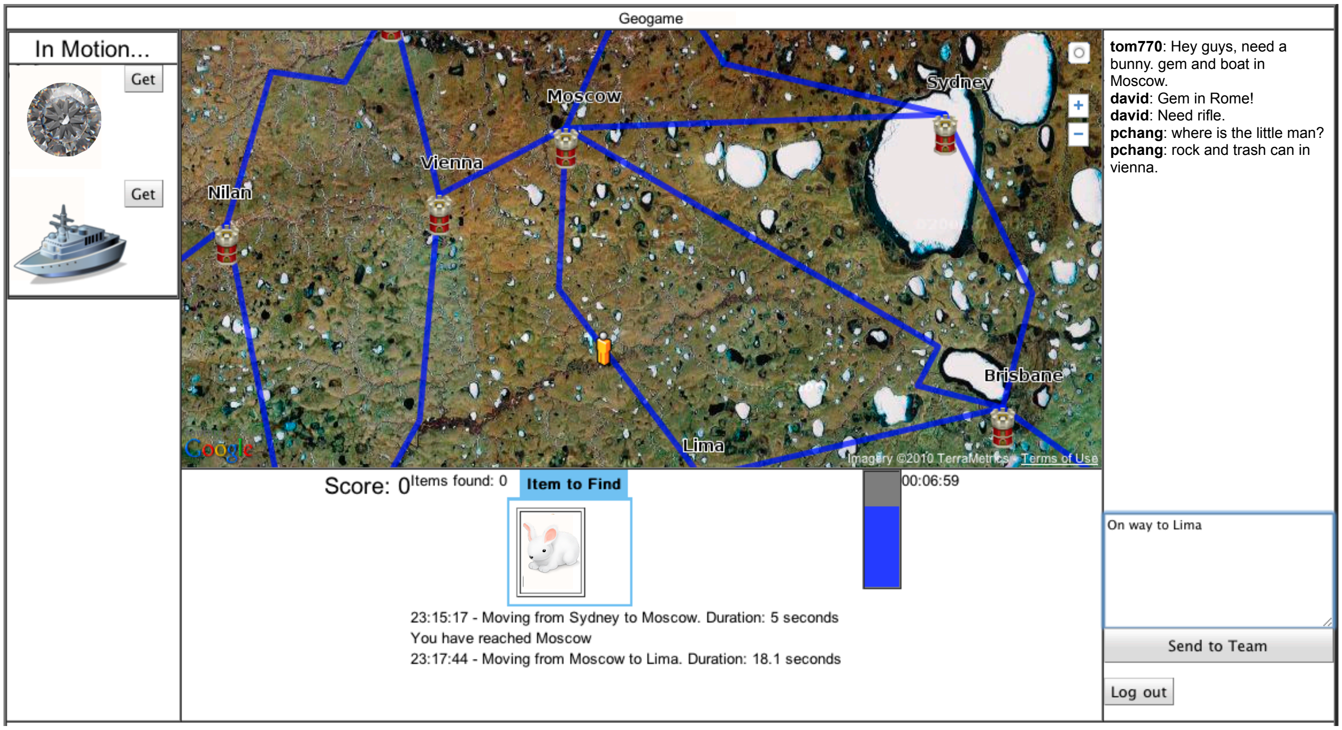

Figure 1. The Geo Game interface for human players during a simulated trial.

On an ontological level, the task operationalizes pieces of knowledge as city-item associations; messages typically contain either requests for the location of an item, or facts. When an item is taken at a location, it is re-created at a different, random location; this implements a dynamic ground truth and invalidates existing facts. The number of items taken is our primary measure of task success. Cooperation and individual efforts were incentivized.

EXPERIMENT

This study is designed to demonstrate a convergence of vocabulary in human communities communicating along the paths defined by small-world networks. Vocabulary convergence means that subject groups will tend to use a common word rather than a diverse set of words to express a given meaning. To show this, we allow participants to choose their words for given images.

Participants were shown unlabeled images that represented different Geo Game target items, as they were available in the cities. Thus, they were free to choose a label such as *gem* or *diamond* to refer to a precious stone during their communications. We can define a *final label* out of such a set of labels for each item by identifying the dominant label in the final portion of the game. Convergence implies that participants use the final label proportionally more often over time.

Four participant groups (average 20 subjects) played four sessions of 20 minutes duration each. Participants were given course credit, but were not remunerated according to their success in the game in order to prevent undue circumvention of the established communication channels. Sessions were closely proctored to avoid communication outside of the system, In sessions 1, 2 and 4, participants communicated according to a randomly generated network with small-world properties (each participant broadcasting to their network neighborhood). In session 3, participants did not communicate at all.

To establish a diverse set of labels, participants underwent a visual priming phase before session 1. There, each participant was shown the set of available items, one at a time, with a label shown underneath the item, e.g., *gem*. Each item displayed for one second, with a one-second blank distractor in between each item. The label was chosen randomly.

I distinguish two types of items: those whose labels are in a nearly synonymous relationship, and those whose labels are in a hyper/hyponym relationship. Examples of near-synonyms include *boat* and *ship*, *rock* and *stone*, or also the non-synonymous but related *cupcake* and *muffin*. I acknowledge that these may not be true synonyms. Examples of hypo/hypernyms include *rifle* and *gun*, or *person* and *man*. Images for the items were selected using both of their labels via Google Image search in order to obtain images that did not exhibit a specific bias for either label.

After each 20-minute session, participants were asked to fill out a post-test recall form, which asked for the items available at each cities (the form provided them with a map and an entry for each city). Thus, they were to recall the items and their labels again at this stage, albeit not in a communicative context.

RESULTS

The written communications between participants were mapped to items via automatic means (spelling variations were ignored, e.g., *bunnie* was interpreted as *bunny*). Data from each 30-minute session and the associated post-test were pooled (an analysis based on time slots did not yield substantially different results).

For each item, all labels are ranked according to their frequency in the final session (4). For each of the other sessions, the dominance of each label can then be expressed in terms of its final frequency rank (1 if it was the item that dominated in the final session).

Over the course of the four sessions, participants adopted a largely common set of label-item associations. Figure 2 shows the odds ratio of choosing label ranked first over the label ranked second. A generalized linear regression model (Table 1) explained adoption of the first-ranked label (as opposed to the second-ranked label - all others excluded) as a function of session number (1,2,4), item type and their interaction (and random effects of session number grouped by item and by subject). A reliable convergence effect was found (adoption over session, $p < 0.005$). The effect remains significant when excluding session 4 (as the ranking was based on this session).

Significant adoption of the primed terms could not be shown (in a different model). However, participants did choose a variety of labels initially. In post-hoc analysis, a stark contrast between convergence during the regular sessions and the label use in the written post-test questionnaires was found. The questionnaires showed no adoption of the joint vocabulary.

To visualize the convergence, an aggregate representation of the languages can be used. Each item i is assigned two labels coded 0,1. I define a vector $\vec{l_{t,s}} = \begin{pmatrix} L_{0,t} \\ \cdots \\ L_{n,t,s} \end{pmatrix}$ encoding the subject s's label choice, $L_{i,t,s}$, for item i around time t. $L_{i,t,s}$ may be binary or continuous, encoding a relative frequency. Thus the nodes in each communication network are, at each point in time, located in a semantic space. The standard cosine metric quantifies similarity between each pair of vectors by measuring their distance in space. A matrix is built encoding the distances between each vector pair. Then, a dimensionality reduction technique is applied (multidimensional scaling in this case) to the matrix to map the semantic space to two-dimensional space for visualization.

Variable	β	SE	$\lvert z \rvert$	p
Intercept	2.214	0.339	6.526	< 0.0001
Session (1-4)	0.442	0.138	3.214	< 0.005
Type ($hyp = -0.5$)	1.048	0.656	1.599	0.11
Session:Type	0.091	0.269	0.339	0.74

Table 1. Binomial regression model predicting adoption of first-ranked label. Random Session variables grouped by item and by subject. All variable values were centered. β coefficients predict response in logit space. (lme4 package in R; p-values by t-test.)

Comparing the resulting network graphs at different time points t yields an animated view of the lexical evolution (Figures 3 and 4). The diagrams show the data from four subject groups, which all used the same set of items and labels, thus a shared semantic space. Several subjects end up with overlapping languages (shown as single circles in the diagram), even though some subjects in each group did not fully converge after the 60 minutes of collaboration.

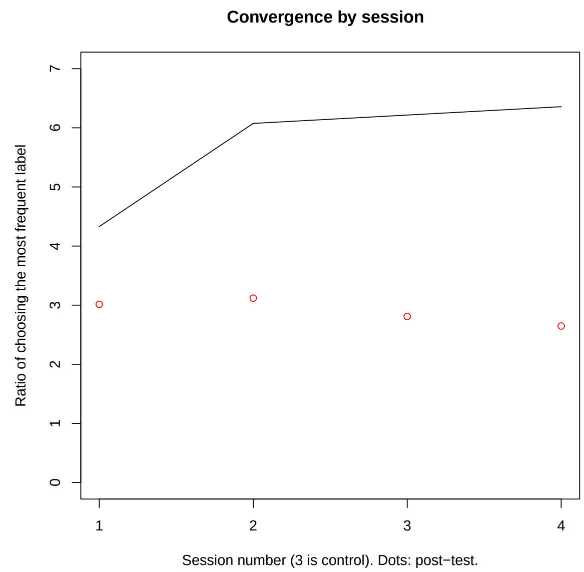

Figure 2. Ratio of most frequent label (as used in session 4) compared to second-most frequent label. Red circles indicate use of labels in the post-test questionnaires.

DISCUSSION AND ONGOING WORK

Language evolution has often been modeled, but only few serious attempts have been made to actually document it in experiments in groups of more than two people. The results show vocabulary convergence, which is a first step to a model that explains and predicts which words will be adopted by a group of users, and how structural properties of the network interact with the convergence process. Such a model would predict the adoption of specific labels as a function of relative prior frequency (*rock* is more common than *stone* for a countable noun), semantic relations between the labels (e.g., item type as above), economic considerations (number of syllables). An evolutionary model may, ultimately, also shed light on the frequencies of labels (even in a large corpus) as a function of the other variables (e.g., it is known that highly frequent words tend to be shorter and morphologically synthetic rather than analytic – *got* rather than *getted* as past tense of *get*).

The observation that participants distinguish between different modalities of communication was somewhat surprising. Models of linguistic adaptation in joint problem-solving [7] state that basic priming (or learning) mechanisms even at the lexical level govern linguistic decision-making. In cognitive models of the learning processes that underlie the adoption of specific linguistic conventions by people, we have em-

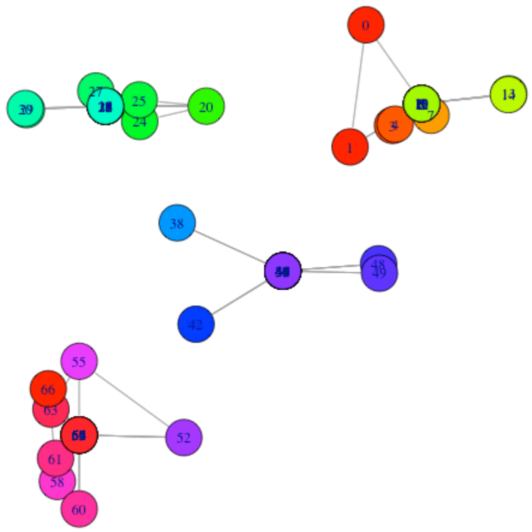

Figure 3. Languages used by four human subject groups during the first 250 seconds (first session). Each vertex in this graph represents the language spoken by one subject. Its location is the result of a mapping from multidimensional space of all possible languages to the two-dimensional space of the graph. Edges represent the communication channels between subjects.

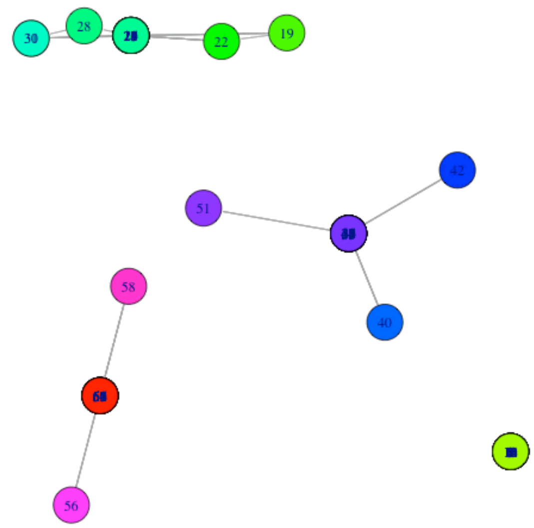

Figure 4. Final 20 minutes of gameplay. More overlap of nodes shows stronger coherence between the languages of a group.

phasized that general-purpose human memory (declarative memory, that is) provides the basis for such convergence processes [8, 9]. The clearly preliminary results suggest that, for the explicit (conscious) process of lexical choice, people do design their message for the specific audience or modality (cf., [5, 3, 2]).

CONCLUSION

The primary contribution of this study is to demonstrate the evolution of lexical choice in groups of human participants. This extends past work in computational modeling, non-verbal (graphical) communication, and also diachronic linguistics. Of course, the fact that vocabularies evolve is hardly surprising. Such results, however, establish a baseline that allows experimenters to determine the variables that influence such processes. The two examples we show are semantic relationships within the ontology forming the basis for the vocabulary, and the role of generic, declarative memory, as participants appear to differentiate between communicative addressees or situations. These form the beginnings of a model that can quantitatively describe which words are adopted, and of presentation techniques that allow us to better visualize such convergence processes. The vocabulary development we have shown clearly takes place in the short term; yet, such rapid adoption can be seen as the precursor to long-term lexical change.

Acknowledgements

The author thanks Christian Lebiere for advice with the Geo Game platform, and Yury Vinokurov for his work in implementing the platform and running experiments. This work was funded by the Air Force Office of Scientific Research (FA 95500810356).

REFERENCES

1. Anderson, J. R. *How can the human mind occur in the physical universe?* Oxford University Press, Oxford, UK, 2007.
2. Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language 42* (2000), 1–22.
3. Clark, H. H. *Using Language*. Cambridge University Press, Cambridge, 1996.
4. Fay, N., Garrod, S., Roberts, L., and Swoboda, N. The interactive evolution of human communication systems. *Cognitive Science 34*, 3 (2010), 351–386.
5. Horton, W. S., and Keysar, B. When do speakers take into account common ground? *Cognition 59* (1996), 91–117.
6. Mason, W., and Suri, S. Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods 44* (2012), 1–23. 10.3758/s13428-011-0124-6.
7. Pickering, M. J., and Garrod, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27*, 2 (2004), 169–190.
8. Reitter, D., Keller, F., and Moore, J. D. A computational cognitive model of syntactic priming. *Cognitive Science 35*, 4 (2011), 587–637.
9. Reitter, D., and Lebiere, C. How groups develop a specialized domain vocabulary: A cognitive multi-agent model. *Cognitive Systems Research 12*, 2 (2011), 175–185.
10. Steels, L. Grounding symbols through evolutionary language games. In *Simulating the Evolution of Language*. Springer, 2001.