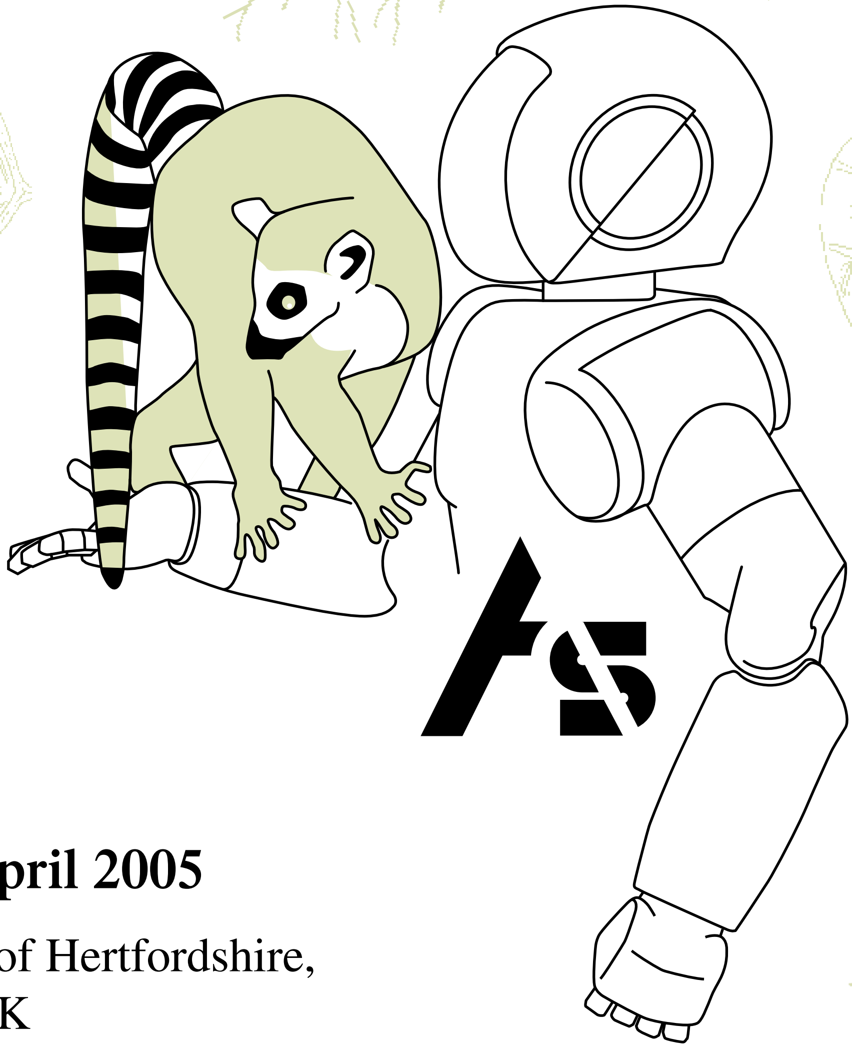
AISB'05: Social Intelligence and Interaction
in Animals, Robots and Agents

Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)



12 - 15 April 2005

University of Hertfordshire,
Hatfield, UK

SSAISB 2005 Convention

EELC

Emergence and Evolution of Linguistic Communication

AISB

UH

EPSRC

Engineering and Physical Sciences
Research Council

AISB'05 Convention

Social Intelligence and Interaction in Animals, Robots and Agents

12-15 April 2005
University of Hertfordshire, Hatfield, UK

Proceedings of

Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)

Published by

AISB

The Society for the Study of Artificial Intelligence and the Simulation of Behaviour
www.aisb.org.uk

Printed by

The University of Hertfordshire, Hatfield, AL10 9AB UK
www.herts.ac.uk

Cover Design by Sue Attwood

ISBN 1 902956 40 9

AISB'05 Hosted by

The Adaptive Systems Research Group
adapsys.feis.herts.ac.uk

The AISB'05 Convention is partially supported by:

EPSRC Engineering and Physical Sciences Research Council

The proceedings of the ten symposia in the AISB’05 Convention are available from SSAISB:

Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)
1 902956 40 9

Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action
1 902956 41 7

Third International Symposium on Imitation in Animals and Artifacts
1 902956 42 5

Robotics, Mechatronics and Animatronics in the Creative and Entertainment Industries and Arts
1 902956 43 3

Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction
1 902956 44 1

Conversational Informatics for Supporting Social Intelligence and Interaction - Situational and Environmental Information Enforcing Involvement in Conversation
1 902956 45 X

Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment
1 902956 46 8

Normative Multi-Agent Systems
1 902956 47 6

Socially Inspired Computing Joint Symposium (Memetic theory in artificial systems & societies, Emerging Artificial Societies, and Engineering with Social Metaphors)
1 902956 48 4

Virtual Social Agents Joint Symposium (Social presence cues for virtual humanoids, Empathic Interaction with Synthetic Characters, Mind-minding Agents)
1 902956 49 2

Table of Contents

The AISB’05 Convention
Social Intelligence and Interaction in Animals, Robots and Agents

Above all, the human animal is social. For an artificially intelligent system, how could it be otherwise?

We stated in our Call for Participation “The AISB’05 convention with the theme *Social Intelligence and Interaction in Animals, Robots and Agents* aims to facilitate the synthesis of new ideas, encourage new insights as well as novel applications, mediate new collaborations, and provide a context for lively and stimulating discussions in this exciting, truly interdisciplinary, and quickly growing research area that touches upon many deep issues regarding the nature of intelligence in human and other animals, and its potential application to robots and other artefacts”.

Why is the theme of Social Intelligence and Interaction interesting to an Artificial Intelligence and Robotics community? We know that intelligence in humans and other animals has many facets and is expressed in a variety of ways in how the individual in its lifetime - or a population on an evolutionary timescale - deals with, adapts to, and co-evolves with the environment. Traditionally, social or emotional intelligence have been considered different from a more problem-solving, often called "rational", oriented view of human intelligence. However, more and more evidence from a variety of different research fields highlights the important role of social, emotional intelligence and interaction across all facets of intelligence in humans.

The Convention theme *Social Intelligence and Interaction in Animals, Robots and Agents* reflects a current trend towards increasingly interdisciplinary approaches that are pushing the boundaries of traditional science and are necessary in order to answer deep questions regarding the social nature of intelligence in humans and other animals, as well as to address the challenge of synthesizing computational agents or robotic artifacts that show aspects of biological social intelligence. Exciting new developments are emerging from collaborations among computer scientists, roboticists, psychologists, sociologists, cognitive scientists, primatologists, ethologists and researchers from other disciplines, e.g. leading to increasingly sophisticated simulation models of socially intelligent agents, or to a new generation of robots that are able to learn from and socially interact with each other or with people. Such interdisciplinary work advances our understanding of social intelligence in nature, and leads to new theories, models, architectures and designs in the domain of Artificial Intelligence and other sciences of the artificial.

New advancements in computer and robotic technology facilitate the emergence of multi-modal "natural" interfaces between computers or robots and people, including embodied conversational agents or robotic pets/assistants/companions that we are increasingly sharing our home and work space with. People tend to create certain relationships with such socially intelligent artifacts, and are even willing to accept them as helpers in healthcare, therapy or rehabilitation. Thus, socially intelligent artifacts are becoming part of our lives, including many desirable as well as possibly undesirable effects, and Artificial Intelligence and Cognitive Science research can play an important role in addressing many of the huge scientific challenges involved. Keeping an open mind towards other disciplines, embracing work from a variety of disciplines studying humans as well as non-human animals, might help us to create artifacts that might not only do their job, but that do their job right.

Thus, the convention hopes to provide a home for state-of-the-art research as well as a discussion forum for innovative ideas and approaches, pushing the frontiers of what is possible and/or desirable in this exciting, growing area.

The feedback to the initial Call for Symposia Proposals was overwhelming. Ten symposia were accepted (ranging from one-day to three-day events), organized by UK, European as well as international experts in the field of Social Intelligence and Interaction.

- Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)
- Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action
- Third International Symposium on Imitation in Animals and Artifacts
- Robotics, Mechatronics and Animatronics in the Creative and Entertainment Industries and Arts
- Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction
- Conversational Informatics for Supporting Social Intelligence and Interaction - Situational and Environmental Information Enforcing Involvement in Conversation
- Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment
- Normative Multi-Agent Systems
- Socially Inspired Computing Joint Symposium (consisting of three themes: Memetic Theory in Artificial Systems & Societies, Emerging Artificial Societies, and Engineering with Social Metaphors)
- Virtual Social Agents Joint Symposium (consisting of three themes: Social Presence Cues for Virtual Humanoids, Empathic Interaction with Synthetic Characters, Mind-minding Agents)

I would like to thank the symposium organizers for their efforts in helping to put together an excellent scientific programme.

In order to complement the programme, five speakers known for pioneering work relevant to the convention theme accepted invitations to present plenary lectures at the convention: Prof. Nigel Gilbert (University of Surrey, UK), Prof. Hiroshi Ishiguro (Osaka University, Japan), Dr. Alison Jolly (University of Sussex, UK), Prof. Luc Steels (VUB, Belgium and Sony, France), and Prof. Jacqueline Nadel (National Centre of Scientific Research, France).

A number of people and groups helped to make this convention possible. First, I would like to thank SSAISB for the opportunity to host the convention under the special theme of *Social Intelligence and Interaction in Animals, Robots and Agents*. The AISB'05 convention is supported in part by a UK EPSRC grant to Prof. Kerstin Dautenhahn and Prof. C. L. Nehaniv. Further support was provided by Prof. Jill Hewitt and the School of Computer Science, as well as the Adaptive Systems Research Group at University of Hertfordshire. I would like to thank the Convention's Vice Chair Prof. Chrystopher L. Nehaniv for his invaluable continuous support during the planning and organization of the convention. Many thanks to the local organizing committee including Dr. René te Boekhorst, Dr. Lola Cañamero and Dr. Daniel Polani. I would like to single out two people who took over major roles in the local organization: Firstly, Johanna Hunt, Research Assistant in the School of Computer Science, who efficiently dealt primarily with the registration process, the AISB'05 website, and the coordination of ten proceedings. The number of convention registrants as well as different symposia by far exceeded our expectations and made this a major effort. Secondly, Bob Guscott, Research Administrator in the Adaptive Systems Research Group, competently and with great enthusiasm dealt with arrangements ranging from room bookings, catering, the organization of the banquet, and many other important elements in the convention. Thanks to Sue Attwood for the beautiful frontcover design. Also, a number of student helpers supported the convention. A great team made this convention possible!

I wish all participants of the AISB’05 convention an enjoyable and very productive time. On returning home, I hope you will take with you some new ideas or inspirations regarding our common goal of understanding social intelligence, and synthesizing artificially intelligent robots and agents. Progress in the field depends on scientific exchange, dialogue and critical evaluations by our peers and the research community, including senior members as well as students who bring in fresh viewpoints. For social animals such as humans, the construction of scientific knowledge can't be otherwise.

Dedication:

I am very confident that the future will bring us increasingly many instances of socially intelligent agents. I am similarly confident that we will see more and more socially intelligent robots sharing our lives. However, I would like to dedicate this convention to those people who fight for the survival of socially intelligent animals and their fellow creatures. What would 'life as it could be' be without 'life as we know it'?

Beppu, Japan.

Kerstin Dautenhahn

Professor of Artificial Intelligence,
General Chair, AISB'05 Convention *Social Intelligence and Interaction in Animals, Robots and Agents*

University of Hertfordshire
College Lane
Hatfield, Herts, AL10 9AB
United Kingdom

Symposium Preface

Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)

EELC

Emergence and Evolution of Linguistic Communication

AISB'05 Convention, 12-15 April 2005, University of Hertfordshire, U.K.

SYMPOSIUM OVERVIEW - EMERGENCE AND EVOLUTION OF LANGUAGE

The renewed scientific interest in the emergence and evolution of linguistic communication has become one of the most important research issues in Artificial Intelligence and Cognitive Science. The EELC'05 Symposium focuses on the latest empirical and modelling research on the evolutionary factors that affect the acquisition, self-organization and origins of linguistic communication systems and their precursors. This considers both language-specific abilities (e.g. speech, semantics and syntax) and other cognitive, sensorimotor and social abilities (e.g. category learning, action and embodiment, social networks). Key questions relate to the emergence of: symbol grounding; deixis, gesture, and reference; predication; negation; syntactic categories; and compositionality; among other issues in the context of embodied, social interaction and evolution. This is a field characterized by a highly interdisciplinary and multi-methodological approach. It benefits from the contribution of researchers from wide ranging disciplines such as linguistics, psychology, neuroscience, anthropology and computer science. The methodologies adopted cover a wide range of approaches, from animal and human experiments, to brain studies and to computational and robotic modelling of linguistic behaviour. For example, computational models of language evolution and emergence involve artificial intelligence methods (e.g. artificial neural networks, evolutionary computation, rule-based systems) and techniques for the simulation of behaviour (artificial life, multi-agent systems, adaptive behaviour and robotics). The symposium creates the opportunity for the many of most influential in the field to present their latest research and to discuss the agenda for future studies.

The use of computational models for simulating the evolution of language has been one of the main contributors to the renewed interest in language evolution research. In fact, up to 10 years ago, very few researchers were directly interested in the origins and evolution of language and publications on new language evolution studies were uncommon. This was partly the result of the famous ban in the 19th century by the Société Linguistique de Paris on research and publication on language origins to quell rampant, unfounded speculation on the topic. The development of the first language evolution models in the early 1990s permitted researchers to deal with some of the main difficulties in such a scientific endeavour. Theories of language origins and evolution not only were difficult to test empirically but they tended to be stated in vague and general terms and were unable to generate detailed empirical predictions. This was partially due to the problem of the scarcity of objective empirical evidence. It is this very problematic aspect of the study of language evolution which computer simulations can help us to overcome. Computer simulations are theories of the empirical phenomena that are simulated (Cangelosi & Parisi 2002). Simulations are a novel way to express theories in science. They are scientific theories expressed as computer programs. The program incorporates a set of hypotheses on the causes, mechanisms, and processes underlying the simulated phenomena and, when the program runs in the computer, the results of the simulations are the empirical predictions derived from the theory incorporated in the simulation. All this contributes to the development of a new approach to the study of the origins and evolution of language.

THE EELC SYMPOSIUM SERIES

Following on from the success of the First International Workshop on the Emergence and Evolution of Linguistic Communication in Japan 2004, and the Evolution of Language conferences, this symposium is held 14-15 April 2005 at the University of Hertfordshire, College Lane Campus, Hatfield, just outside London. It is part of the AISB 2005 Convention 12-15 April 2005, whose overall theme is "Social Intelligence and Interaction in Animals, Robots and Agents". EELC'04 was the First International Workshop on the Emergence and Evolution of Linguistic Communication (EELC), held in Kanazawa (Japan) in May/June 2004 under the auspices of the Japanese Society for Artificial Intelligence (JSAI), the Japanese counterpart of AISB, at the JSAI 2004 Convention. The Second EELC Symposium, now at AISB'05 in the U.K., aims to continue the philosophy of this meeting and its international tradition. This is particularly relevant since both British and Japanese scientists have played a major role on the development of computational models of language evolution. In addition, the location of the workshop within the AISB annual meeting permits a better exchange with other researchers working in the field of artificial intelligence and simulation of behaviour, whether they work in Britain or come from abroad to attend the meeting.

The aims of the EELC symposium are:

- to provide an common interdisciplinary forum for researchers of the emergence and evolution of language,
- to discuss and disseminate the latest research on theoretical, empirical and modelling investigations of the evolution of linguistic communication and its precursors,
- to set the agenda for future research and identify the most promising theoretical and methodological issues in the area.

Acknowledgements. The symposium was supported in part by a grant of the British Academy, the National Academy for Humanities and Social Sciences. We also thank the programme committee and local organizing committee of EELC’05, and the AISB 2005 Convention organizers, for their dedicated work, as well as all the authors and speakers in making this symposium a success!

Angelo Cangelosi and Chrystopher L. Nehaniv
Programme Chairs, EELC'05

EELC'05 SYMPOSIUM ORGANIZATION

Invited Speakers:
Luc Steels (AI Lab Vrije Universiteit Brussel, Belgium)
Alison Wray (Cardiff University, Wales)
W. Tecumseh Fitch (University of St. Andrews, Scotland)

Programme Chairs:
Angelo Cangelosi (University of Plymouth), Chair
Chrystopher L. Nehaniv (University of Herfordshire), Co-Chair

Local Organizing Committee
Caroline Lyon (University of Hertfordshire), Local Organizing Committee Chair
Jean Baillie (University of Hertfordshire), Local Organizer
Gianluca Massera (University of Plymouth), Local Organizer

International Scientific Programme Committee
Takaya Arita (University of Nagoya, Japan)
Jean Baillie (University of Hertfordshire, UK)
Aude Billard (EPFL, Switzerland)
Angelo Cangelosi (University of Plymouth, UK)
Takashi Hashimoto (JAIST, Japan)
Koiti Hasida (AIST, Japan)
Jim Hurford (University of Edinburgh, UK)
Takashi Ikegami (University of Tokyo, Japan)
Simon Kirby (University of Edinburgh, UK)
Caroline Lyon (University of Hertfordshire, UK)
Chrystopher L. Nehaniv (University of Hertfordshire, UK)
Stefano Nolfi (ISTC, National Research Council, Italy)
Kazuo Okanoya (University of Chiba, Japan)
Tetsuo Ono (Future University Hakodate, Japan)
Domenico Parisi (ISTC, National Research Council, Italy)
Akito Sakurai (Keio University, Japan)
Luc Steels (Vrije Universiteit Brussel, Belgium)
Satoshi Tojo (JAIST, Japan)

Colourful language and colour categories

Tony Belpaeme[1,2] Joris Bleys[1]

[1] Artificial Intelligence Lab
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
{tony,jorisb}@arti.vub.ac.be

[2] University of Plymouth
School of Computing, Communication and Electronics
Portland Square, Plymouth, PL4 8AA, United Kingdom

Abstract

We investigate whether the universal character of colour categories can be explained as the result of a category acquisition process under influence of linguistic communication. A brief overview is presented of the different positions in explaining the mechanisms of colour category acquisition (or perceptual categories in general). We introduce a computational model to study the acquisition of colour categories with and without linguistic interactions. We present preliminary results, which are compared with recent results from the World Color Survey. We argue that combining biases from colour perception, perceptual categorisation and linguistic communication provides an alternative explanation for the nature of colour categories.

1 Introduction

For more than three centuries the precise nature of human colour categories has been one the most disputed topics among physicists, psychologists, cognitive scientists and anthropologists. Newton, already in the eighteenth century, wondering about the number of categories that could be discerned in the sunlight's spectrum, decided on the divine number of seven, thereby requiring a category called "indigo" that no-one had observed until then. Three centuries later much more precise data on colour categories is available and together with the data came a plethora of interpretations.

One of the most influential contributions is the monograph by Berlin and Kay (1969) in which they reported on the linguistic colour categories of 20 languages. Using naming experiments they elicited the colour categories of subjects and comparing the categories across different languages they noticed a remarkable cross-cultural correspondence. Until then the general consensus had been that colour categories were random for each culture, but Berlin and Kay's work rekindled the conviction that the universal character of colour categories could only be explained as being genetically determined.

In this paper we first summarise the results of the World Color Survey (WCS) (Kay et al., 1997, 2003) reported in (Kay and Regier, 2003). This work provides the strongest evidence yet of strong universal tendencies in colour naming in seperate languages. We give an overview of the different accounts which try to explain this universal character and then continue to present a computational model which tests whether linguistic relativism might be a viable candidate. We report several results from the simulation and compare these with the data from the WCS.

2 The World Color Survey

The WCS reports on colour naming experiments with speakers of 110 languages spoken in non-industrialised societies. The field data has been gathered in North and South America, Africa and South-East Asia.

In the study each subject is shown a series of 330 coloured chips drawn from the Munsell colour set [1]

[1] The Munsell Color Company (GretagMacbeth, New Windsor, NY) produces calibrated chips for art reproduction. The most saturated chips have been used by anthropologists to study colour categories since the 1950s.

(of which 320 chips show gradations of hues at different lightness, all at maximal saturation, and 10 chips show shades of grey, ranging from white to black) and asked to name each chip.

The analysis of the data proceeded as follows. For all subjects studied, the centroid was computed for every colour term they used. For this the Munsell colour values were converted to the CIE $L^*a^*b^*$ colour appearance model[2]. The term centroids were projected back onto the closest matching Munsell chip. For each language a chart can now be produced showing the average representation of all colour terms in that language.

To get a visual impression of the linguistic colour categories over all 110 languages, the centroids of all subjects of all languages can be combined into one single histogram (figure 1). The floor plane of the histogram corresponds to the ordered Munsell chart, with on one axis the hue value of the chip, ranging from red, over yellow, green, blue, to purple; and on the other axis the lightness of the chip (note that it does not display the counts for achromatic chips).

The histogram shows that the linguistic colour categories of different languages are not arbitrary; it clearly illustrates the universal character of colour categorisation. Peaks can be found at regions close to the English colour terms pink/red, brown, yellow, green, blue and purple.

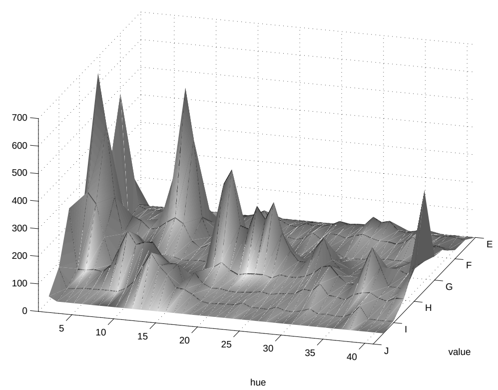


Figure 1: Histogram showing the linguistic colour categories for 110 languages spoken in non-industrialised societies (data from Kay and Regier, 2003).

[2]CIE $L^*a^*b^*$ is a perceptually uniform colour representation. It is a 3D colour space, in which the L^* dimension represents the lightness, and the a^* and b^* dimensions represent the chroma of the colour. A Euclidean distance function can be used to compute the perceptual distance between two CIE $L^*a^*b^*$ values.

3 Attempts at explaining universalism

The challenge now is accounting for colour naming universalism. The leading position has always been that colour categorisation results directly or indirectly from an innate endowment (Kay and McDaniel, 1978; Bornstein, 1985; Hardin, 1988; Shepard, 1992; Kaiser and Boynton, 1996). One hypothesis states that there exist basic colour categories that are explicitly related to the opponent colour processing in the human visual pathways. Psychological and neurophysiological data indeed points to an opponent character of human colour perception, with white contrasting with black, red with green and yellow with blue. All other basic categories —orange, brown, pink, purple and grey— can be deduced from these six primaries. Although this account has made it into textbooks (e.g. Crystal, 1997), some scholars still doubt that colour categories are unequivocally fixed by neural correlates (Saunders and van Brakel, 1997; Lucy, 1997; Jameson and D'Andrade, 1997) or that colour categories are universal at all (Roberson et al., 2000).

In the next section we will present a computational model to study if colour categories can be explained as a concept formation process which is under influence of language (or cultural exchange in general). It has been proposed by some that colour categories not only are associated with colour terms, but that colour terms also have an influence on the acquisition of colour categories (Gellatly, 1995; Davies and Corbett, 1997). This position has become known as the Sapir-Whorf hypothesis (Whorf, 1956).

4 The computational model

The computational model we use is based on a research methodology first proposed by Steels (1996a,b). Using this methodology Steels studied how meanings can be associated unambigiously with words. It was later extended for studying adaptive meanings and open lexica in (Steels, 1998; Belpaeme, 2001). The methodology relies on multi-agent simulations. Each agent is able to perceive, categorise its perceptions and lexicalise the resulting categories. We briefly present the internals of an agent:

Perception The perception of colours is modelled by relying on the properties of the CIE $L^*a^*b^*$ colour space (Fairchild, 1998). Agents are offered colour stimuli as RGB triplets, these are converted to CIE $L^*a^*b^*$ values. The conversion from RGB to CIE $L^*a^*b^*$ is given in the following equations. The conversion matrix is for PAL/SECAM viewing conditions, with $\gamma = 2.5$; the XYZ coordinates of the reference white are taken to be $[X_n Y_n X_n]^T = [0.950\ 1.000\ 1.089]^T$.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.431 & 0.222 & 0.0202 \\ 0.342 & 0.707 & 0.130 \\ 0.178 & 0.0713 & 0.939 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}^{\gamma}$$

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right) - 16 & \frac{Y}{Y_n} > \epsilon \\ 903.3\left(\frac{Y}{Y_n}\right) & \frac{Y}{Y_n} \leq \epsilon \end{cases}$$
$$a^* = 500\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Y}{Y_n}\right)\right)$$
$$b^* = 200\left(\mathrm{f}\left(\frac{X}{X_n}\right) - \mathrm{f}\left(\frac{Z}{Z_n}\right)\right)$$
$$\mathrm{f}(x) = \begin{cases} x^{1/3} & x > \epsilon \\ 7.787x + 16/116 & x \leq \epsilon \end{cases}$$
$$\epsilon = 0.008856$$

The CIE $L^*a^*b^*$ colour representation was designed to mimick human psychological colour experience, and therefore serves well as our colour perception model.

Categorisation To implement perceptual categorisation we resort to a point representation in the CIE $L^*a^*b^*$ space. Each colour category is a point in that space, and the membership function for a category is the Euclidean distance to that point.

Lexicalisation Colour categories can be associated with colour terms. The strength of the association is represented by a scalar value $s \in [0, 1]$. Colour categories can be associated with more than one word (thereby allowing synonymy) and words can be associated with more than one category (thereby allowing homonymy).

Additionally an interaction between two agents is defined, which serves to let the agents acquire a repertoire of colour categories and colour terms. The interaction implements horizontal transmission of lexical entries and categories. It consists of two components, a *discrimination game* and a *guessing game*, both described below.

4.1 The discrimination game

The discrimination game serves to build a repertoire of categories that allows an agent A to distinguish between stimuli. This pseudo code for the discrimination game is as follows.

Algorithm 1 Discrimination Game(A, $\mathcal{O}$)

1: Agent A chooses a topic o_t from the context $\mathcal{O} = \{o_1, \ldots, o_N\}$ containing N objects.
2: Agent A perceives each stimulus in the context by constructing an internal representation for it: $\{o_1, \ldots, o_{\mathcal{O}_N}\} \rightarrow \{r_1, \ldots, r_N\}$
3: For each internal representation r_i, the best matching category is found. This is the category which has the highest output for r_i of all the categories available in the category repertoire of the agent $A_{\mathcal{CR}}$ and which we will denote by c_i^{best}: $\{r_1, \ldots, r_N\} \rightarrow \{c_1^{best}, \ldots, c_N^{best}\}$
4: If the best matching category for the topic c_t^{best} is unique in $\{c_1^{best}, \ldots, c_{\mathcal{O}_N}^{best}\}$ the game succeeded, otherwise it has failed.

An agent is offered a number of objects, this is called the *context* $\mathcal{O}$. One of the object is the *topic* o_t, which the agent has to distinguish from the other objects in the context. For this, the agent first perceives all objects, which results in a number of *internal representation* r_i. Next, the internal representation are matched to categories. For example, if the agent has only one category, all representation of objects will be matched to that same category, making it impossible for the agent to distinguish betweem objects. However, as soon as the agent has more than one category, it can start distinguishing between objects. If the topic is matched with a category with which no other object matches, we say that the agent is able to "discriminate the topic from the context" and we call the discrimination game a success.

The discrimination game can fail in several ways: this is an opportunity to improve the agent's categorical repertoire. When the category repertoire $A_{\mathcal{CR}}$ is empty, a new category is created on the internal representation of the topic r_t. When no discriminating category could be found, there are two possible actions: (1) a new category is created on r_t or (2) the best matching category c_t^{best} is adapted to better represent the internal representation of the topic r_t, this is done by shifting c_t^{best} towards r_t. Option (1) is taken when the discriminative success of the agent is below a threshold $\theta_{adapt} = 0.95$, otherwise option (2) is taken.

4.2 The guessing game

The guessing game is played between two agents randomly chosen from the population: one acting as *speaker* (A_S) and the other as *hearer* (A_H). The pseudo code for the guessing game is as follows [3].

The speaker and hearer both observe the same context $\mathcal{O}$. The speaker knows what the topic o_t of the conversation is, and tries to linguistically communicate the topic to the hearer. For this the speaker first plays a discrimination game, if this succeeds the speaker looks up the word associated with the discriminating category. This word is then relayed to the hearer. The hearer looks up the category belonging to the word, and maps the category onto the objects in the context. It then points to the object which matches best with the category. Finally, the speaker reports if the hearer has pointed correctly to the topic. During the course of the guessing game, both agents adapt the strength s_{ij} between category c_i and word w_j according to the following equation (with $\delta = 0.1$).

$$\begin{cases} s_{ij} = \min\left(s_{ij} + \delta, 1\right) \\ s_{kl} = \max\left(s_{kl} - \delta, 0\right) \\ \quad\quad \text{in row i and column j with } k \neq i, l \neq j \end{cases} \tag{1}$$

A categories is adapted by shifting the point representation of a category towards a representation r, as in eq. 2; α is a learning rate, set to 0.7.

$$c \leftarrow c + \alpha(r - c) \tag{2}$$

Of course, also the guessing game can fail at several ways. For each failure, an appropriate action is taken so that the agents will be more successful at communicating in future games.

- The speaker fails at the discrimination game: it adapts its categorical repertoire as described in 4.1.
- The speaker has no word associated with c_t^{best}: a new word is created and associated with an initial strength $s = 0.5$.
- The hearer does not know the word w: the speaker "points" at the topic and the hearer associates the word w with the category best matching the topic, with initial strength $s = 0.5$.
- The hearer fails to pick out the topic ($o_t \neq o_h$): the strength of the association between c_t^{best} and w is decreased by δ.

[3]DG stands for discrimination game.

When the guessing game is successful the speaker and hearer both increase the strength of the association between the categories used and the communicated word[4].

5 Experimental results

As input to the agents we use two different sets of colour data. One set, called the *random* set, contains random colours generated by drawing colours from the RGB colour solid and then converting them to CIE $L^*a^*b^*$. The other set, called the *nature* set, draws colours from digital photographs of natural scenes. The difference between both is that the *random* set contains a uniform distribution of colours, while the *nature* set contains a skewed distribution with an abundance of low-saturated colours and few high-saturated colours. The purpose of having two data sets is to study the effect of the environment on the acquisition of colour categories.

For reference the results from the WCS (Kay and Regier, 2003) are repeated in figure 2 now a contour plot of figure 1. The locations of English colour terms are added for reference.

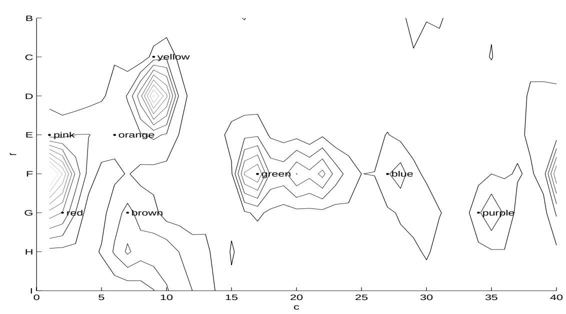


Figure 2: Contour plot of the WCS data.

Four types of simulations have been run. DGRAN: discrimination game where agents are fed *random* data. DGNAT: discrimination game where agents are fed *nature* data. GGRAN: guessing game where agents are fed *random* data. And GGNAT: guessing game where agents are fed *nature* data.

Each type of simulation has a population of 10 agents and has been run 105 times[5]. The results presented for each type of simulation are the sum of these 105 runs.

[4]More details, specifically on the implementation of the update rules, can be found in (Bleys, 2004; Steels and Belpaeme, 2005)

[5]One could think of these 105 runs as hundred different artificial societies.

Algorithm 2 Guessing Game(A_S, A_H, O)

speaker A_S		hearer A_H
chooses topic o_t		
plays DG for o_t		
DG succeeds and returns c^S		
finds term w for c^S		
utters w	$\rightarrow w \rightarrow$	hears w
		finds category c^H for w
		finds o_h closest to c^H
sees o_h	$\leftarrow o_h \leftarrow$	points to o_h
if hearer guessed right, then $o_t = o_h$		
update s^S_{cw} using eq. 1		
points to o_t	$\rightarrow o_t \rightarrow$	sees o_t
		updates s^H_{cw} using eq. 1
		adapts category c^H to r_t using eq. 2

Figures 3, 4, 5 and 6 show contour plots of histograms collecting the colour categories of 10×105 agents. A first observation is that each type of simulation cuts up the colour continuum in a number of peaks: colour categories are not randomly constructed (if they would be, the histogram should not have any peaks).

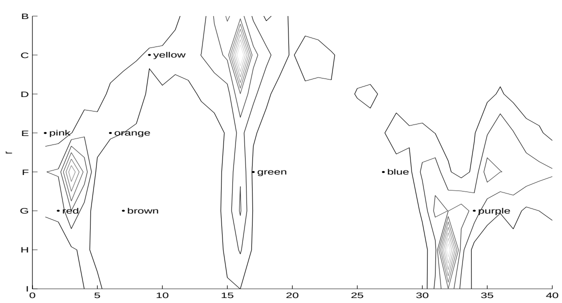


Figure 3: Contour plot of DGRAN results.

Two biases, present in all four simulations, are quite influential. On the one hand, the psychological colour space —modelled by the CIE $L^*a^*b^*$ colour space— puts constraints in the location of the categories (the colour space is shaped like two bumpy cones connected to each other at their base). The second bias is formed by the property of categories to be maximally distinctive. Both biases act together so that colour categories are in a way "pushed" towards locations where they are maximally distinctive and where they form a stable configuration. Colour categories are stable when they are located in places where shifting the colour category would result in a lower discriminative or communicative success.

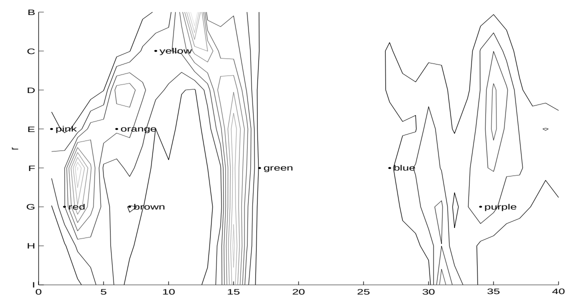


Figure 4: Contour plot of DGNAT results.

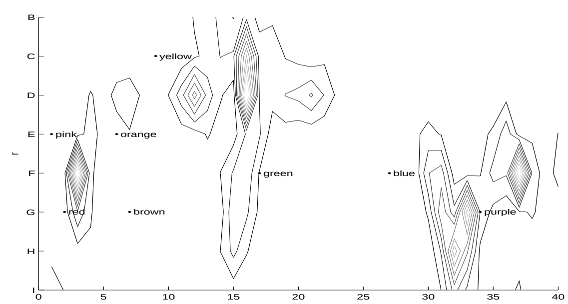


Figure 5: Contour plot of GGRAN results.

In this sense, all four simulations return colour categories that retain all properties of human perceptual categories. However, the purpose of our study is to see whether acquiring colour categories with an additional bias formed by linguistic communication would result in categories that are more human-like. Figures 5 and 6 when compared to figure 2 give a

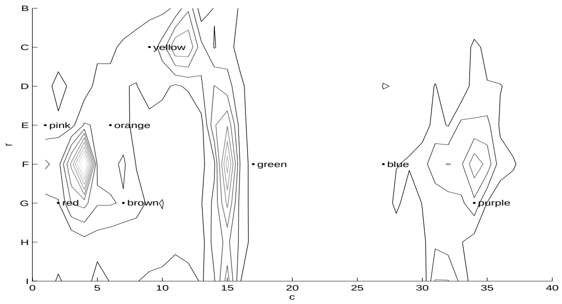


Figure 6: Contour plot of GGNAT results.

qualitative impression, but a measure is needed compare the histograms quantitavily. Eq. 3 computes the sum of squared pair-wise differences between two histograms h and h' (with h_i being the bin at index i in histogram h).

$$d(h, h') = \sum_i (h_i - h'_i)^2 \qquad (3)$$

Table 1 shows the comparison between the histograms obtained from the simulations and the WCS data. According to the measure we use, the DGNAT simulation resembles human colour categories most. However, also the DGRAN and GGNAT data have a similar distance to the human data. Only the GGRAN data seems to be off, why remains eludes us at the moment.

$d(h, WCS)$	
DGRAN	0.00973
DGNAT	0.00842
GGRAN	0.0145
GGNAT	0.00994
WCS	0

Table 1: Sum of squared differences between simulation histograms and WCS data (lower values correspond to a higher similarity).

6 Discussion

The computational models that are presented here implement a view on colour categorisation which contrasts with the innatist viewpoint on colour categories. We have shown how agents can acquire a set of categories that is sufficient to discriminate colours, and in the case of the guessing game simulations, the agent acquire colour categories that not only discriminate well, but also communicate well.

The categories resulting from the simulations are qualitatively similar to human colour categories: they take up regions in the colour space that correspond well to the WCS data. We have not been able to show that the influence of communication on category formation results in radically different categories. This might however be due to the limitations of our analysis. The sum of squared distances measure might not be suited to compare two-dimensional histograms. For example, if two identical histograms are compared, but one is shifted relative to the other, the sum of squared distances measure will return a low value; this is not desired.

Future analysis will point out if there exist measures which might give a better impression of the similarity of histograms. One alternatively could be to extract the peaks of the histograms and compare the using a certain distance measure[6].

Acknowledgements

The authors wish to thank Terry Regier and Paul Kay for the helpful discussion and for making the World Color Survey data available. Tony Belpaeme is a postdoctoral fellow of the Flemish Fund for Scientific Research (FWO Vlaanderen). The research reported was in part sponsored by the FWO Verbal Communication on Humanoid Robots project.

References

Tony Belpaeme. Simulating the formation of color categories. In Bernhard Nebel, editor, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 393–398, Seattle, WA, 2001. Morgan Kaufmann, San Francisco, CA.

Tony Belpaeme and Joris Bleys. Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior*, 2005. Submitted.

Brent Berlin and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA, 1969.

Joris Bleys. The cultural propagation of colour categories: Insights from computational modelling. Master's thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, 2004.

[6]This is further explored in (Belpaeme and Bleys, 2005)

Marc H. Bornstein. On the development of color naming in young children. *Brain and Language*, 26:72–93, 1985.

David Crystal. *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge, MA, 2nd edition, 1997.

Ian R.L. Davies and G.G. Corbett. A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, 88:493–517, 1997.

Mark D. Fairchild. *Color Appearance Models*. Addison-Wesley, Reading, MA, 1998.

Angus Gellatly. Colourful Whorfian ideas: Linguistic and cultural influences on the perception and cognition of colour, and on the investigation of them. *Mind and Language*, 10(3):199–225, 1995.

Clyde L. Hardin. *Color for Philosophers: Unweaving the Rainbow*. Hacket, Indianapolis, 1988.

Kimberley Jameson and Roy D'Andrade. It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C.L. Hardin and L. Maffi, editors, *Color categories in thought and language*, pages 295–319. Cambridge University Press, Cambridge, 1997.

Peter M. Kaiser and Robert M. Boynton. *Human Color Vision*. Optical Society of America, Washington DC, 1996.

Paul Kay, Brent Berlin, Luisa Maffi, and William Merrifield. Color naming across languages. In C.L. Hardin and L. Maffi, editors, *Color Categories in Thought and Language*. Cambridge University Press, Cambridge, 1997.

Paul Kay, Brent Berlin, Luisa Maffi, and William R. Merrifield. *The World Color Survey*. Center for the Study of Language and Information, Stanford, 2003. ISBN 1575864150.

Paul Kay and Chad McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, 54(3):610–646, 1978.

Paul Kay and Terry Regier. Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15):9085–9089, 2003.

John A. Lucy. The linguistics of "color". In Clyde L. Hardin and Luisa Maffi, editors, *Color Categories in Thought and Language*, pages 320–346. Cambridge University Press, Cambridge, 1997.

Debi Roberson, Ian Davies, and Jules Davidoff. Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3):369–398, 2000.

Barbara A.C. Saunders and Jaap van Brakel. Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20(2):167–228, 1997.

Roger N. Shepard. The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. Barkow, L. Cosmides, and J. Tooby, editors, *Adapted Mind*, pages 495–532. Oxford University Press, Oxford, 1992.

Luc Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multiagent Systems (ICMAS-96)*, pages 338–344, Menlo Park, CA, 1996a. AAAI Press.

Luc Steels. Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Proceedings of the Conference on Artificial Life V (Alife V) (Nara, Japan)*, Cambridge, MA, 1996b. The MIT Press.

Luc Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1:169–194, 1998.

Luc Steels and Tony Belpaeme. Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences*, 2005. Accepted as target article. Available from http://arti.vub.ac.be/˜tony.

Benjamin Lee Whorf. *Language, Thought and Reality: selected writings of Benjamin Lee Whorf*. The MIT Press, Cambridge, MA, 1956. Edited by Carrol, J.B.

Self-organized Diversification of Signals of Different Species

Alexander Bisler
*Fraunhofer Institute for Computer Graphics
Darmstadt, Germany
abisler@igd.fhg.de

Abstract

This paper reports results of a computer simulation which models the behavior of different species of communicative agents which share the same habitat. The agents use signals to communicate. This communication is in a male/female context: males use their signals to attract females. Since both share the same habitat, the signals of the species have to be distinguishable, allowing females to identify a male of her own species. But signals used by males of the same species should be similar. The simulation shows that these conventions can emerge using mechanisms of self-organization.

1 Introduction

Research in the origins and evolution of language has shown that a population of communicative agents can reach a shared set of conventions without a global control (Steels, 1996a). Language is now seen as a complex adaptive system which becomes coherent through mechanisms of self-organization.

The self-organized formation of vocabularies has been studied by Ke et al. (Ke et al., 2002) and Steels (Steels, 1996b). A set of agents starts with randomly created vocabularies. But the agents must share vocabularies to achieve cooperative benefits through communication. Under certain conditions a coherent vocabulary can emerge.

On the other hand, some research dealt with modeling linguistic diversity (Livingstone and Fyfe, 1999; Arita and Koyama, 1998), where linguistic stochasticity is a source for innovation.

In further scenarios, diversity of signals is strongly intended as, for example, when two or more bird species share the same habitat. Male birds use their songs to attract females[1] (Davies; Vehrencamp, 2000; Attenborough, 1998). But since females of other species hear these songs, too, it is vital that a convention be established to distinguish the different species. A female approaching a male of another species would just waste time and energy.

This paper presents a simulation of how different signals for different species can emerge from the interactions of simple communicative agents. All species start using the same signal. Since these signals are not distinguishable for the females, the males start to modify their signals so that each species uses signals which are distinguishable from the other ones. But this, of course, leads to different signals within a species. So each species has to make the effort to establish a common signal again. All this is driven by the need to attract females of the same species.

[1] They also use them to defend their territory and alarm others.

2 Simulation Setup

In each simulation run there are m populations of communicative agents of m different species. Each population has a specific size. The ratio of male to female birds is controlled by a simulation parameter.

The agents use signals S to communicate. A signal $s \in S$ is modelled with an array of scalars[2]. The nth element of the array is denoted as s_n. A randomly created signal is composed of sine-waves of different amplitudes.

Two signals x and y are similar if the squared difference does not exceed a specific limit l:

$$\sum_{i=1}^{n}(x_i - y_i)^2 \leq l$$

In nature male birds can use a variety of different songs by rearranging different parts or modifying them. In this simulation only a simple signal is used to show the basic idea of communication between agents.

[2] The size of the array was set to 60.

3 Actions & Attributes of Agents

An agent can fly around freely. Male birds can emit signals. All agents, which are in the near vicinity of the source of such a signal, receive it.

Each agent has a memory where it can store signals. The agents stores one specific signal s_{own} which is associated with their own species. Male birds emit this signal to attract females, and female birds use this signal to compare it with those heard previously as an aid in finding males of their own species.

Each agent also stores a set of signals $S_{foreign}$ which is associated with other species. Females hearing a signal which is similar to those of $S_{foreign}$ do not react to it. Signals in $S_{foreign}$ tend to fade out over time, unless they are heard again and can be associated with a different species. Therefore, they have a weight associated with them which is initialized by a value F, decreased each simulation frame by an offset f_{-}, and increased by an offset f_{+} if it can be associated with a different species. When the weight reaches 0, the signal is removed from the memory.

When male birds think that their signal s_{own} is not successful enough, they can either create a new one or modify it. Failed signals are stored in a set $S_{failures}$ which has only a limited capacity $\#S_{failures}$. If it is full, the signal, which was stored first, is deleted.

To imitate other males of the same species, they can adapt their signal step by step.

The male birds also have a value which defines their dominance. Female birds tend to mate with more dominant males.

When female birds enter the scenario, they do not possess a signal s_{own}. When they meet the first male of their species, they remember his signal. When mating with other males, they gradually adapt their signal to the signals used by these males.

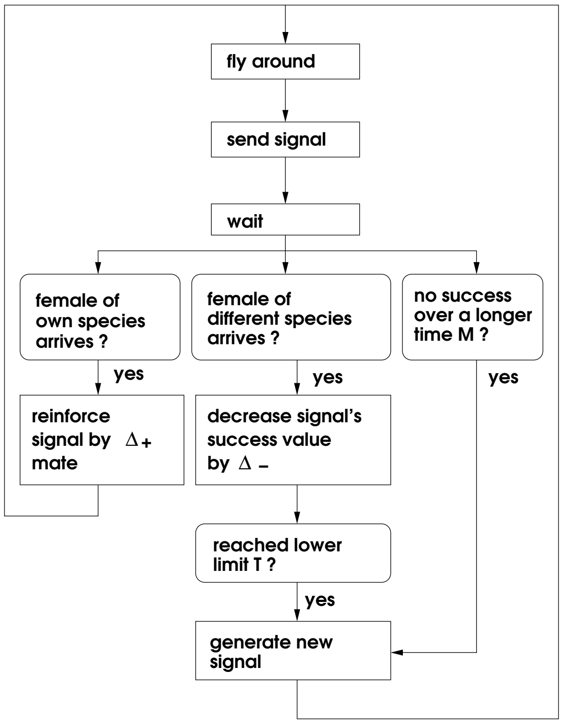


Figure 1: Flow diagram of male behavior.

4 Guidelines

The experiments follow the guidelines first set out by Hutchins and Hazlehurst (Hutchins and Hazlehurst (1995), quotations from Curran and O'Riordan (2002)).

- "The limited rationality constraint: No mind may influence another except via a mediating structure (no telepathy)." – Agents react only to signals they received. Signals are of an acoustic nature. They can only be received within a limited listening range of the source.
- "The distributed system constraint: No agent has a complete view of all other agents." – An agent can only see and hear other agents which are in its closer vicinity.
- "The open system constraint: There should be an in-flux/out-flux of agents throughout simulations to examine the stability of an emerging communication system."– A steady replacement of agents with new ones was established. The new agents came into the system with randomly generated signals. The rate of replacements was controlled.
- "No social mind can become organized except via interaction with another or its environment." – The only chance for an agent to become organized is to interact with other males (imitating their signals if they belong to the same species) or females (rewarding his signal if he attracts a female or modifying it otherwise).
- "The nature of a mental representation cannot simply be assumed, it must be explained." – An agent stores the received signals in its memory. It associates these signals with its own species or another one. Signals of foreign species are forgotten after some time in case they are not

heard anymore. Additionally, own signals are associated with a success value.

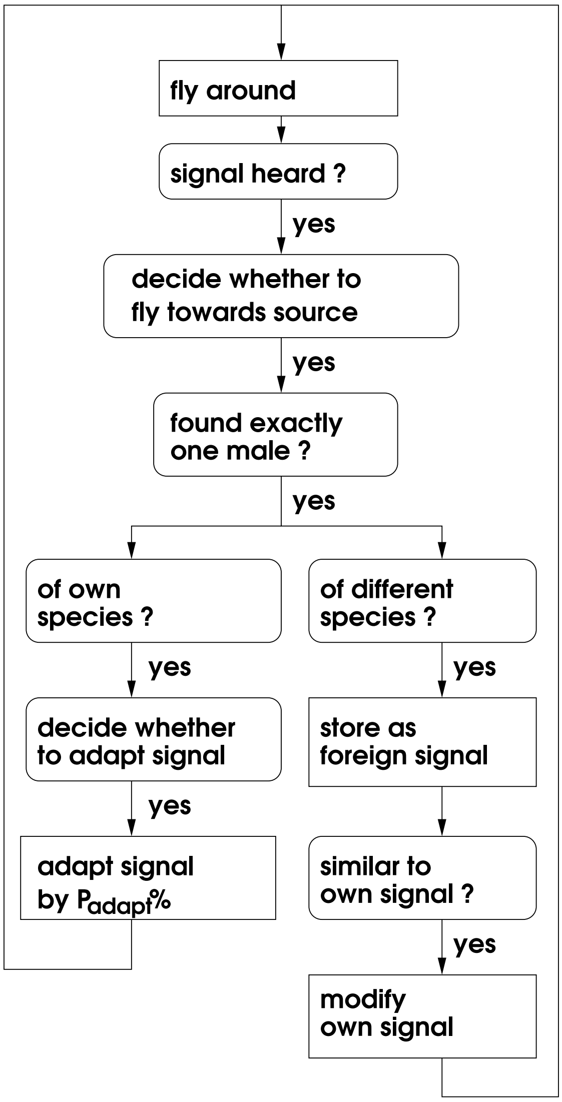


Figure 2: Flow diagram of male imitation behavior.

5 Male Behavior

A male bird flies around and periodically emits a signal (i.e. sings a song). He then waits for a response. If no female bird arrives, he moves on (Figure 1).

If a female of his species arrives which wants to mate, his signal is rewarded by an offset Δ_+. If a female of another species arrives, he decreases the success value of his signal by an offset Δ_-. If this value reaches a lower limit T, the male bird generates a new signal. The old one is remembered as not successful and stored in $S_{failures}$. The agent ensures that the new signal is significantly different from his other unsuccessful trials in $S_{failures}$ and from foreign signals in $S_{foreign}$. If no females arrive at all within a certain time M, he stores his signal in $S_{failures}$ and creates a new one.

When a male bird hears a new signal, he can also decide to follow the signal to its source and explore to which species it belongs (Figure 2). When he arrives at the source location and sees exactly one male, he assumes that he is the source of that signal. It is possible that the emitter has already flown away and a male of another species has arrived at that location. This leads to a false assumption. But since signals in $S_{foreign}$ tend to fade out, this mistake is not permanently retained.

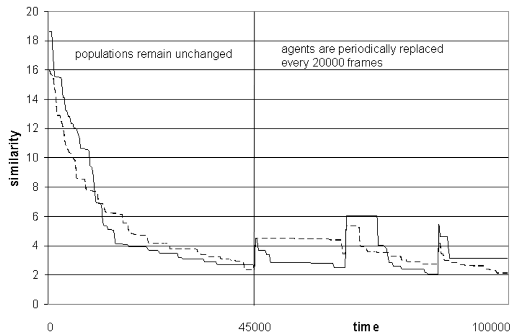


Figure 3: Similarity of signals within a species (shown for two species). The lower the value, the smaller the differences. In the closed system a commonly used signal emerges. Then, the system is opened and remains stable.

If the observed male is from the same species and more dominant, the male bird decides to adapt his own signal to the newly heard one (by $P_{adapt}\%$). If the observed male is less dominant, the probability that the more dominant one adapts to the heard signal is less likely.

If the observed male is from another species, the signal is stored in $S_{foreign}$. If the agent's own signal is similar to the one that he heard, he decides to modify it so as to be distinguishable from the other species.

Then the agent flies in the direction where he observes fewer males. Otherwise male agents tend to occupy a single location chatting with each other.

If the number of successful contacts is less than the number of contacts with females of a different species, the male tends to follow more often an unknown signal. If he is more successful, he sees no need in doing this and decides more likely against it.

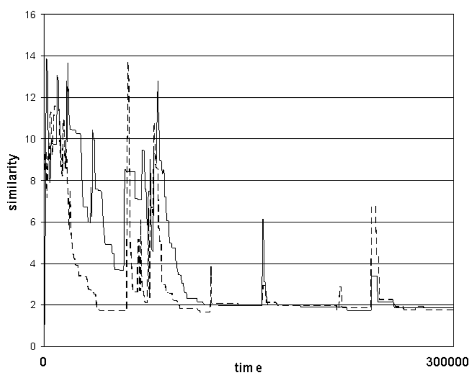


Figure 4: Similarity of signals within a species (A_2: straight line, B_2: dashed line). The lower the value, the more similar the signals are.

6 Synchronization

In a first scenario with two different species (40 agents each) but without female birds, the male behavior was observed. The agents all started with different signals. The simulation measured the variety of signals within each species.

At first the system was closed, i.e. no agent left or entered the scenario. The development of the signals showed that the variety was reduced. After each species converged to a commonly accepted signal, the system was opened and agents were periodically replaced with new ones. Figure 3 shows the development of the similarity of the signals. The lower the value, the smaller the differences. It shows that the convention of a signal can emerge in a closed society and remains stable when new agents enter the system afterwards. This is comparable to the results Steels found in his experiments (Fig. 3 in Steels and Kaplan (1998)). If the exchange rate is too high, the system becomes unstable and diverges again.

7 Female Behavior

Female birds also fly around and listen to signals. When they hear one, they compare it to those they already associate with other species, stored in $S_{foreign}$ (Figure 14). If it is similar to those, they just ignore it. But if it is similar to s_{own} or if they have not learned

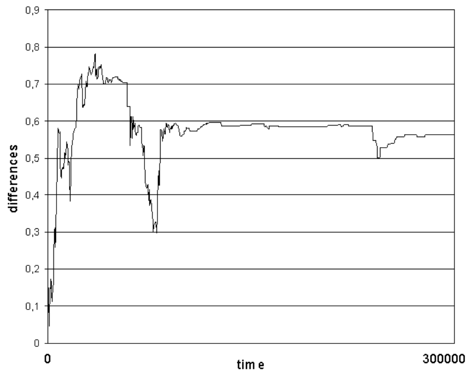


Figure 5: Differences of average signals between species A_2 and B_2. The larger the value, the larger the difference.

any signal at all, they fly towards its source location. If the signal is unknown, the probability to follow it is reduced.

If they follow the signal, they search for a male. If they find exactly one, they assume that he is the emitter. Here, the same source of errors as for males exists.

If the observed male belongs to the same species, the female decides whether to mate with him or not. The more dominant a male is, the more likely the female decides to mate. If the female has not learned a signal so far, she takes this one as s_{own}. Otherwise, she adapts her signal s_{own} to the signal of the male (by $P_{adapt}\%$). This means, that females gradually adapt to changes in male signals. Additionally, she increases the success value of s_{own} by the offset Δ_+.

If the observed male belongs to another species, the female bird compares the heard signal to s_{own}. If s_{own} is similar, she decreases its success value by the offset Δ_-. If the lower limit T is reached, s_{own} is stored in $S_{foreign}$. If it is not similar to s_{own} , she stores the heard signal in $S_{foreign}$.

8 Examination of the Parameters

The simulation consisted of a number of parameters. First, a configuration of this parameter set was examined by manually modifying the values. This provided some insights within what ranges the values led to a successful solution and why other configurations failed.

To examine a greater number of possible configurations, the parameter settings were examined with a genetic algorithm (Holland, 1975; Goldberg, 1989). Therefore, the parameters were divided into two groups. Those which described the "physical reality" of the environment and those which influenced the behavior of the birds. The physical parameters could not be modified by the genetic algorithm. Only the behavior parameters were put into a genome:

- Probability that dominant males adapt their signal to less dominant males.
- Percentage of the most dominant males which are always chosen by females.
- Offset Δ_+ by which the success value of s_{own} is increased.
- Offset Δ_- by which the success value of s_{own} is decreased.
- Lower limit T for failed signals. If the success value of s_{own} falls below this limit, it is exchanged by a newly created signal and stored in $S_{failures}$.
- Maximal number $\#S_{failures}$ of failed signals which a male can remember.
- Initial weight F for foreign signals.
- Offset f_- by which a foreign signal's weight is decreased each frame.
- Offset f_+ by which a foreign signal's weight is increased if recognized again.
- Percentage P_{adapt} by which a bird adapts its signal to the signals of other birds of its species.
- Length M of time interval while a male bird remembers how many females wanted to mate with him.
- Number of contacts a male remembers. For each contact he remembers also whether it was a success or not. If the number of failures is greater than the number of successes, the probability to move towards the source of an unknown signal is set to a higher value (60%). Otherwise it is set to a lower value (30%).
- Ratio of females to males. By decreasing this ratio the probability of contacts is decreased forcing the males to adapt their behavior.

As the fitness function the success rate of all females of all species was used. The population size[3] was set to 45 and the search lasted for 200 generations[4]. The probability of mutation was set to 30% and the probability for crossover to 80%.

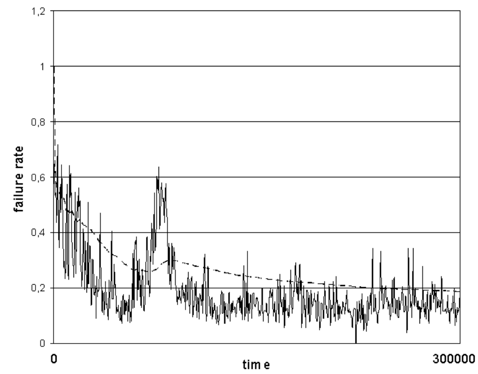


Figure 6: Rate of female failures (of both species A_2 and B_2): average of all attempts (dashed, falling line) and average of attempts within the last 500 frames (jagged line).

9 Diversification

In a first scenario two species A_2 and B_2 were set into the same habitat. Each population consisted of 40 agents. All male birds of all species started with the same signal. The system was open, so that agents were periodically replaced with new ones every 80000 frames. The whole simulation lasted for 300000 frames.

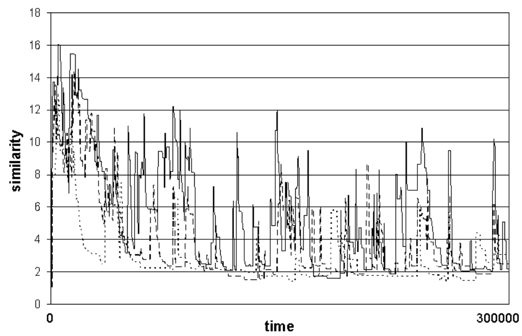


Figure 7: Similarity of signals within each species. The lower the value, the smaller the differences (A_3 straight line, B_3 dashed, C_3 dotted).

[3] the number of simulations running parallel

[4] for all scenarios

In a second scenario three species A_3, B_3, and C_3 were set into the same habitat. Each population consisted of 30 agents.

And in a third scenario four species A_4, B_4, C_4, and D_4 were set into the same habitat. Each population consisted of 25 agents.

The parameter configurations for all scenarios were examined using the genetic algorithm described above. The scenario with two species was also examined by manually modifying single parameters.

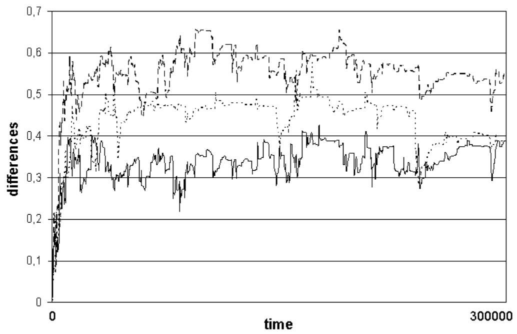


Figure 8: Differences of average signals between species A_3 and B_3 (straight), A_3 and C_3 (dashed), and B_3 and C_3 (dotted). The larger the value, the larger the difference.

10 Results

10.1 Two Species

Figure 4 shows the development of the similarity of the signals used within each species. In the beginning the similarity tended to worsen in a very short time. The initial signals were too close to those of the other species as they either attracted too many females of the wrong species or no females at all as they discovered that these signals were not to be trusted.

But after the first phase of diversification, the males of each species started to synchronize their signals and the similarity improved, while the distinction between the two species increased (Figure 5).

The successful and unsuccessful attempts of the females to approach males of their own species were also measured. Figure 6 shows the ratio of failures to all attempts: the dotted line shows the ratio for all females of all species; the thin (jagged) line represents the ratio for the last 500 frames. The lines start at 1.0, representing total failure, but drop quickly showing the success of the newly learned signals.

After 100000 frames the range of the jagged line becomes smaller and smaller and the absolute values are also falling below 0.2 which means that less than 20% of the attempts were failures.

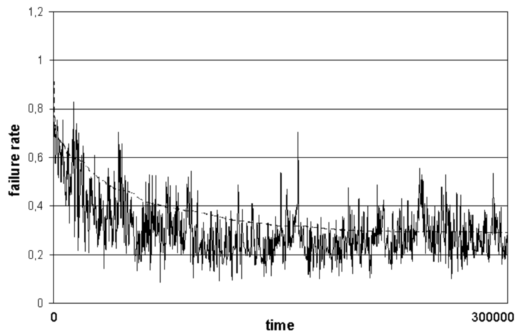


Figure 9: Rate of female failures (of all species A_3, B_3, C_3): average of all attempts (dashed, falling line) and average of attempts within the last 300 frames (jagged line).

The same scenario with a doubled population size showed similar results. But as the size is further increased the probability of reaching a convention decreases as Ke et al. (Ke et al., 2002) and Steels (Steels, 1996b) already have observed.

In early test runs of the diversification scenario the ratio of females to males was kept at one to one. This led to too many contacts between males and females. No evolutionary pressure was built up to force the males to change their signals. Therefore, the number of females was reduced to decrease the probability of a contact.

Another important factor was the rate with which agents were replaced in the habitat. If this replacement rate is too high, a population can not reach a convention about a common signal.

The diversification scenario was also initiated with randomly generated signals for all males. This induced better starting conditions and the failure rate of the females dropped faster, while the inter-species diversification was reached faster as was the inner-species synchronization.

10.2 Three Species

In the second diversification scenario, three species were set into the habitat. The diversification and the synchronization took much more time. Figure 7

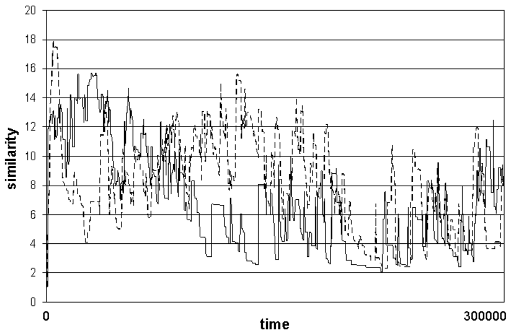


Figure 10: Similarity of signals within each species. The lower the value, the smaller the differences (A_4 straight line, B_4 dashed).

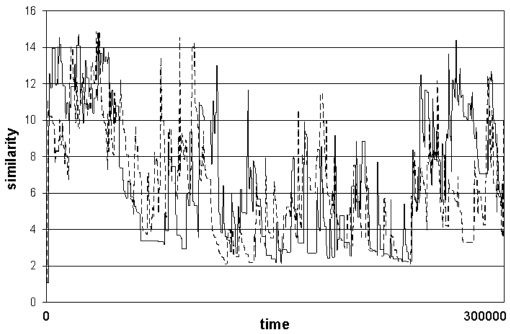


Figure 11: Similarity of signals within each species. The lower the value, the smaller the differences (C_4 straight line, D_4 dashed).

shows the development of the similarity of the signals used within each of the three species. Similar to the scenario with two species, the males of each species started to synchronize their signals and the similarity improved after a phase of diversification in the beginning. But the system was not as stable as in the other scenario. Especially species A_3 (straight line) had difficulties to find a common signal. Species C_3 (dotted line) was most successful.

This can also be seen in Figure 8 which shows the differences of the average signals between species A_3 and B_3 (straight line), A_3 and C_3 (dashed line), and B_3 and C_3 (dotted line). The differences of C_3 to A_3 and B_3 are the largest.

The rate of female failures of all species, as shown in Figure 9, decreased over time, but stayed in the range between 20% and 40%. Thus, the females of all three species were never as successful as those in the scenario with only two species.

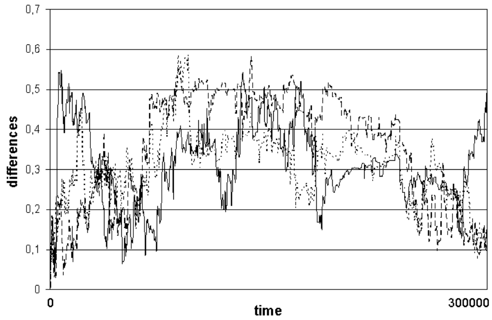


Figure 12: Differences of average signals between species A_4 and B_4 (straight), A_4 and C_4 (dashed), and A_4 and D_4 (dotted). The larger the value, the larger the difference.

10.3 Four Species

In the final diversification scenario, four species were set into the habitat. The system could not come to a solution as in the other two scenarios. Figure 10 shows the similarity of signals for species A_4 and B_4, Figure 11 for C_4 and D_4. The similarities improved from time to time, but they never stabilized.

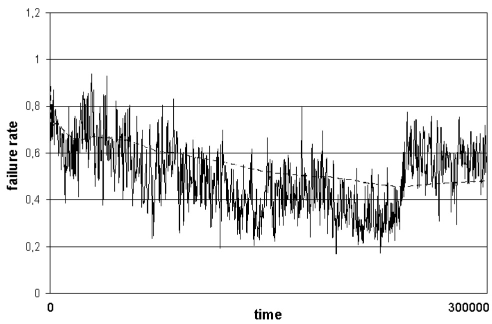


Figure 13: Rate of female failures (of species A_4, B_4, C_4, D_4): average of all attempts (dashed, falling line) and average of attempts within the last 300 frames (thin line).

Also the differences between the species could not be increased to a certain level and remain there. Figure 12 shows the differences of the average signals of

A_4 to all other species. After nearly 200000 frames the emerged differences decreased again.

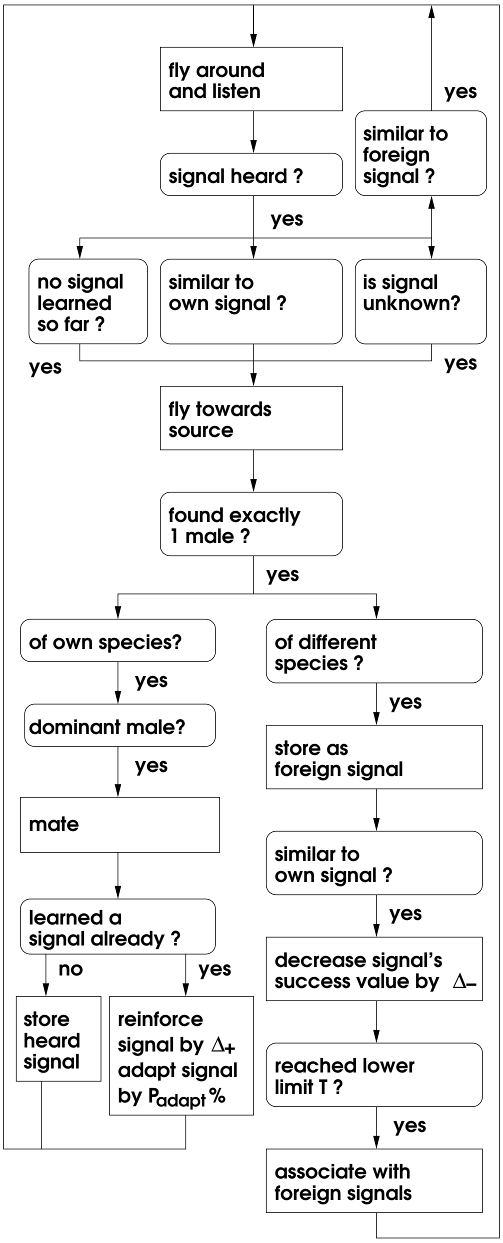


Figure 14: Flow diagram of female behavior.

The rate of female failures of all species (Figure 13) also failed to improve constantly. After 240000 frames the number of failures suddenly increased again.

11 Conclusion

The communicative agents were capable of memorizing signals, associating them with their own or other species, generating, adapting, and comparing them. Additionally, they could measure their mating success. Based on only these basic possibilities and simple rules the different interacting groups of agents were able to organize their behavior.

A positive feedback loop caused some signals to proliferate and eventually become dominant within a species. A negative feedback loop led to diversification of signals between species. The results were obtained without any global control, but only by simple interactions of the agents.

Within a limited range, the system found a solution which was favorable for all species. Changing specific parameters led to an unstable systems. This way the conditions could be identified which were important for a successful solution.

The system never reached a perfect state, since there was always some in- and out-flux and finding the correct emitter of a signal was not ensured. But as soon as the convention about signals starts to be passed on genetically – from generation to generation – better results might be possible.

References

Takaya Arita and Yuhji Koyama. Evolution of linguistic diversity in a simple communication system. In C. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of Artificial Life VI*, pages 9–17. MIT Press, 1998.

David Attenborough. *The Life of Birds*. Princeton University Press, 1998. ISBN 069101633X.

Dara Curran and Colm O'Riordan. Language evolution in artificial systems. Technical report, Dept. of IT., NUI, Galway, 2002.

Gareth Huw Davies. Bird songs. URL `http://www.pbs.org/lifeofbirds/songs/index.html`.

David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201157675.

John Henry Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.

E. Hutchins and B. Hazlehurst. How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte, editors, *Artificial Societies: The Computer Simulation of Social Life*, pages 157–189. UCL Press: London, 1995.

Jinyun Ke, James W. Minett, C.-P. Au, and William S-Y. Wang. Self-organization and selection in the emergence of vocabulary. *Complexity*, 7(3):41–54, 2002.

Daniel Livingstone and Colin Fyfe. Modelling the evolution of linguistic diversity. In D. Floreano, J. Nicoud, and F. Mondada, editors, *Advances in Artificial Life: Fifth European Conference on Artificial Life*, pages 704–708. Springer-Verlag, 1999.

Luc Steels. Emergent adaptive lexicons. In P. Maes, M. J. Mataric, J. A. Meyer, J. Pollack, and S. W. Wilson, editors, *Fourth International Conference on Simulation of Adaptive Behavior; Cape Cod, Massachusetts*, pages 562–567. MIT Press, 1996a.

Luc Steels. Self-organising vocabularies. In Christopher G. Langton and Taksunori Shimohara, editors, *Proceedings of Artificial Life V*. MIT Press, 1996b.

Luc Steels and Frédéric Kaplan. Stochasticity as a source of innovation in language games. In C. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of Artificial Life VI*, pages 368–378. MIT Press, 1998.

Sandra L. Vehrencamp. Handicap, index, and conventional signal elements of bird song. In Y. Espmark, T. Amundsen, and G. Rosenqvist, editors, *Animal Signals: Signalling and Signal Design in Animal Communication*, pages 277–300. Tapir Academic Press, 2000.

Why Information can be Free

Ivana Čače[⋆†]
[⋆]Cognitive Artificial Intelligence
Department of Philosophy
University of Utrecht
The Netherlands
ivana@phil.uu.nl

Joanna J. Bryson[†]
[†]Artificial models of natural Intelligence (AmonI)
Department of Computer Science
University of Bath
Bath, BA2 7AY, UK
jjb@cs.bath.ac.uk

Abstract

This paper describes a model that demonstrates that sharing knowledge can be adaptive purely for its own sake. This is despite the fact that sharing knowledge costs the speaker in terms of foraging opportunities, and that initially the majority of the population consists of free-riders who listen but do not speak. The population is able to take advantage of the increased carrying capacity of the environment that results from the spread of knowledge, and the free riders are reliably out competed by the speakers.

1 Introduction

1.1 Is Language Costly Signalling?

Many contemporary theories on the emergence of linguistic communication focus on mechanisms that could compensate for the cost of communication to the individual. When transmitted information is viewed as a commodity rather then as a means of influencing the behaviour of the receiver, the evolutionary benefits of linguistic communication do become unclear. The cost for a knowledge transmitter is the cost of giving up a competitive advantage by not keeping valuable information to themselves.

Dessalles (2000) uses agent based modelling to argue against the sharing of information as reciprocal altruism. The starting point of his argument is that the gathering of information is costly and that therefore passing it to others for free is effectively costly too. Dessalles' claims that the results of running his agent model where there is a cost to communicating show that communication cannot be a form of reciprocal altruism under realistic conditions. Only if information were very valuable, the cost for dispensing it very low and if it were easy to detect non-cooperative individuals, would the model favour communication. He then proceeds to explain the benefits of language as a means of forming coalitions and obtaining status within a social group through coalitions.

We will not argue against this last point — Dessalles makes a strong case for how language can serve this purpose. Nor will we argue against any other possible benefits of language that have been brought forward over the years, for example its utility in thinking (e.g. Dennett, 1996).

We do challenge the hypothesis that compulsively sharing valuable knowledge is in itself not adaptive. We have built a model in which agents communicate about food. Agents only have access to food resources they know how to exploit and being told is one way of obtaining this knowledge. The cost of dispersing information is that it enables other, competing agents in the vicinity to gain access to limited resources that otherwise may well have been monopolised by the original knowledgeable, communicating individual. Although there is this individual cost, the population as a whole benefits from the dispersal of knowledge as it effectively increases the carrying capacity of the environment by opening up additional food resources. Our model shows that there are circumstances in which communicating agents outcompete silent, knowledge-hoarding agents, despite the fact that these silent agents hear and exploit all the same knowledge the communicating agents in their vicinity hear and exploit.

In the remainder of this section we give a brief introduction in Agent Based Modelling. In the next section we provide a description of our model, then we report the results of the experiments more finally. We conclude with a discussion and summary of results.

1.2 Agent Based Modelling

Agent Based Modelling (ABM) is used as a research tool in a variety of areas including: social science and economics (Epstein and Axtell, 1996; Cederman and Rao, 2001), animal behaviour (Hemelrijk, 2000) and complex systems in general (Baray, 1998; Esteva et al., 2001). In ABM, only the agents and their environment are modelled explicitly. The global behaviours being studied in the system emerge from local interactions, both between agents and between agents and their environment, when the model is run. For example, in our models the agents only communicate to their neighbours who are also their immediate competitors for food. This means that for the individual there is an immediate disadvantage in communicating. However, the global effect we show is that over a longer period of time and for the entire population, communication can be beneficial, and will in some circumstances be selected.

Our models were built in NetLogo (Wilensky, 1999). NetLogo is a freely available multi-agent modelling environment, specifically designed for the ABM of natural and social phenomena. The NetLogo world consists of two kinds of programmable agents: an environment divided into patches and mobile agents called turtles. Patches can represent environmental change, such as growing food, while turtles are typically used to represent animals, including humans.

2 Method

2.1 Prerequisites

In this section we describe our model's implementation. We begin with an overview of the general properties of the model such as the relation between the amount of food and the population size. Then we explain how the environment is built and finally we describe agent behaviour.

The model we use has to be suitable for simulating evolutionary pressure and to model the effects of the spread of knowledge through communication. The first requirement entails that the model yields a more or less stable population so that it can be studied over a long period of time. In this context 'stable' means that the population does not become extinct and does not become so small that chance factors can kill off otherwise successful minority species, but also does not explode in size since this would slow the simulation down a good deal.

For the same reason large, rapid fluctuations in population size are undesirable. As a result, the model has to be initiated with values close to the equilibrium. The consequences of events that cause the population to drop dramatically and the ability (and the speed) to get back to a stable population are of course very interesting, but fall outside the scope of the current research.

In the simulation, time is measured in cycles, one time-step corresponds to one cycle in which all the agents have been activated. So activation is done in a drawing without replacement fashion. Choosing the order in which to activate the agents is resolved by NetLogo itself.

2.2 Basics, the Environment and Food

The number of agents is determined by their birth and death rate, and these in turn are both affected by the replacement rate of the food. The agents reproduce asexually and the reproduction preconditions are chosen such to meet the above-mentioned stability requirement. The offspring function is discussed later in this section, but depends on energy levels. Agents die if they run out of energy or reach maximum age. A maximum age of 50 cycles is imposed to keep a small number of long-lived individuals from influencing the spread of knowledge. In addition, to live the agents need to keep their energy level above zero. They do this by feeding.

Model runs are begun with the values for agent energy and the number of agents close to equilibrium for the given amount of food, including special food. Runs begin with a brief period of population / food oscillation, but these damp quickly to initial equilibrium levels. The population equilibrium then rises slowly as knowledge spreads through the agents, while the food equilibrium stays roughly constant.

The environment consists of 201x201 patch square on a torus space, which is presented on screen as a square. This means agents that walk off any 'edge' of the square will reappear on the opposite edge. On every cycle, energy available in the environment is supplemented in the environment by 'growing food' in a method similar to Wilensky (1998). For every patch, a random number is generated; if it is smaller than the food replacement rate, food is added to the patch.

There are two categories of food: regular food that is available to all agents and six different types of special food with twice as much nutritional value that are only accessible to agents with the corresponding know-how, which enables that food-type's exploitation[1]. A patch can contain only one unit of food. To

[1]This idea was derived from one due to Steele (2004), which in turn derived from the Expensive-Tissue Hypothesis (Aiello and Wheeler, 1995). Steele's idea was that language and a larger brain may have co-evolved. Communication enabled agents to exploit richer food sources which in turn allowed them to have larger brains and smaller guts, thus conserving overall metabolic cost.

prevent the special food from clogging up the environment if no agents know how to eat it, special foods can be overgrown by the regular food using the following algorithm.

On each cycle, food growth is accomplished by generating a random number n for all empty patches, and filling those with an n smaller then the replacement rate for the special food with some special food. Then this procedure is repeated for all the empty patches and all patches containing special food using the replacement rate for regular food. This time, if the random number for a patch that is either empty or containing special food is smaller than the replacement rate for regular food, the patch is emptied and refilled with the regular food.

3 Knowledge and Transmission

There are two breeds of agents, agents that communicate and agents that don't. All agents of both breeds understand communication, and will use knowledge received to eat special food if they find it. The only differences between the breeds is that, every time cycle, agents of the communicating breed choose one food type they know about and communicate this knowledge to all their neighbours.

New knowledge comes into the simulated population in a slightly unrealistic way. Agents are all born knowing how to eat the most basic, lower calorie food type. In addition, when the agents are born they have a 5% probability of knowing how to eat any one of the different, special, food types. No learning except from communication is done after the agent is born. Although somewhat unrealistic, this is relatively easy to code. Further it simulates the propensity of young agents to engage in exploration. Imagine if you like that during their first time step of life, agents are more likely to put strange things into their mouths. Again, both breeds are equally likely to acquire new knowledge this way.

The probability of an individual having offspring depends on their energy level. To keep the population size from fluctuating too much we have chosen a conservative offspring function. Even agents with relatively high energy levels do not necessarily reproduce; rather reproduction is probabilistic, though the probability increases with the energy level. When agents with high energy do reproduce, their offspring are at an advantage compared to that of agents with less energy because at birth the parent's energy is split 80:20 with its offspring.

The offspring is always of the same breed as the parent. This may seem too deterministic, but it is easy to see how a mutation probability biases the ratio between the two breeds. Assume for instance we have a population of 100 individuals of 2 breeds, a 1:10 ratio and a mutation probability of 10% (for explanatory purposes, this is of course too high to be realistic). If the rest of the parameters are such that the population and the ratio between the breeds should stay the same — all individuals hatch once and die immediately afterwards — this is what happens the first time around:

minority breed at $t + 1 = 10 - 0.1 * 10 + 0.1 * 90 = 10 - 1 + 9 = 18$
majority breed at $t + 1 = 90 - 0.1 * 90 + 0.1 * 10 = 90 - 9 + 1 = 82$

Because we want communication to be the only force influencing evolutionary pressure such a bias would be undesirable.

The agents move across their environment randomly with step lengths distributed according to a Levi-flight distribution. Walking patterns fitting this description have been found in foraging animals as well as in evolutionary optimised foraging agents (van Dartel et al., 2002). The formula describing a Levi-flight distribution is:

$$P(l) = 1/z * 1/l^m$$

Where z is a normalising constant and m is a value between 1 and 3. In our implementation we deviate a little from Levi-flight proper by taking $1/m$ to be 3. This is done to keep the knowledge distribution more localised by keeping step length relatively small. Notice though that the density of the population has no impact on their mobility: any number of agents can be standing in the same patch.

The agents lose a small amount of energy with every time step regardless of the distance travelled. Because agents wander around aimlessly they are presented with feeding opportunities at random, but their chances of actually feeding depend on two things:

1. the know-how they possess, and
2. how many neighbouring agents have the same know-how and may thus be eating the same local resources.

3.1 General Experiment Characteristics

The models are run for a long period of time, typically 12,000 time-steps, corresponding to roughly 350 generations. During this time a number of values are recorded and plotted against the duration of the simulation and recorded separately: the number of agents, the ratio between the two breeds, the amount of food in the environment, the amount of regular food and the total amount of the special food. The number of things each agent knows, summed up and divided by the total number of agents, is taken as measure of know-how in the population and is also recorded and plotted. While the values above describe properties of the entire population, some properties of individual agents are also recorded. Namely, when an agent dies its ID, age, date of birth, breed, know-how, parent and all offspring are recorded to a file.

Every experiment started out with a population of 10% communicating agents and 90% silent agents. Also, for each experiment, we ran a control version where all conditions were the same except the 'communicating agents' did not actually communicate. They were however still tagged as a different breed from the normal/silent agents.

3.2 Metrics for Evaluation

Recording values as described above provides the means to measure the influence of communication and the opportunity to examine the mechanisms behind it. Values taken from individual agents and related to breed, food distribution and know-how include the following frequencies:

1. Number of offspring.
2. Number of offspring that managed to reproduce.
3. Age at death.

Biological fitness is nothing but reproductive success. Having many offspring is only an indication of fitness if those offspring get to reproduce. For this reason both 1) and 2) are taken as a metric. The age at death is used by Baray (1998) to measure the effectiveness of cooperative behaviour. We use it as an extra metric; it is useful because it correlates with other population properties like offspring survival rate.

There are also two systemic measures: the breed ratios and the environmental carrying capacity. The latter is simply the average number of agents the environment can sustain — Epstein and Axtell (1996) use this to measure the influence of trade. In the models presented here the amount of food in the environment is determined from the start of every experiment, but the amount of food available to the agents depends on the spread of knowledge. With a greater spread of knowledge more food will become accessible, changing the carrying capacity of the environment.

The breed ratios provide a straightforward way of determining how some trait influences fitness. It simply requires checking if individuals possessing that trait take over the population. It is not a sufficient measure because populations can sometimes become extinct despite being well adapted, especially in a model where the total number of individuals lower than about 2000. Because there are fairly significant but essentially random fluctuations of the population, small populations will die out, even if they might ultimately have proved more adaptive. For this reason, if less than 1000 agents are in the initial population, the initial (less than 100) speakers often die out early in the simulation.

4 Description of Runs

The growth rate of the food provides a handle for manipulating population density and consequently the amount of communication. This holds primarily for the growth rate of the regular food, the growth rate of the special food does not have much effect on the population until the appropriate know-how has percolated through the population. Most of the spread of knowledge happens simultaneously with the communicating agents taking over the population; perfect knowledge on the other hand never seems to be achieved.

In order to start the simulation with reasonable values, the amount of special food in the environment has to be approximately the same as the amount of food when the simulation reaches equilibrium. Recall that the equilibrium value depends on the replacement rate and energetic value of the food and on agent life properties. The latter include the offspring function, maximum age and the amount of energy agents lose every cycle. In the first series of experiments the effects of the different food ratios were measured by keeping the growth rate of the regular food fixed and varying the growth rate of the special food. Remember that the 'special food' differs from the regular food by having a twice as high energetic value and a more limited availability.

4.1 Preliminary Results

We ran experiments with different amounts of food in the environment. Under almost all of the conditions the silent agents die out. As expected, with comparatively more special food in the environment the talking agents had more of an advantage. The time at which all the non-communicating agents die out is inversely proportional to the replacement rate of the special food. Only when the amount of special food in the environment is kept very low does the advantage of communicating disappear. This is done by

food replace rate		average population size	
regular	*special*	*with comm.*	*without comm.*
16	1	1367.51	1336.21
16	2	1440.35	1373.81
16	8	2303.71	1447.97
16	16	4267.45	1579.35

Table 1: The average population size after equilibrium is achieved in two different conditions (with and without communication) at four different replacement rates for special food.

setting a (very low) limit to the amount of food, the food only gets replenished if the sum total falls under this threshold.

With communication the carrying capacity of the environment increases with the amount of special food. In the communication-free control conditions there is also a population increase but much smaller. This increase is due to the individuals who are randomly born knowing about special food, even though they can't communicate about it. Table 1 shows the total population size by the end of the simulation (set to 12,000 time cycles, which is usually well after a stable equilibrium is found.) Notice that in the last two cases the silent agents were already extinct at the time the population stabilised enough to get an average, but in the other cases the population stabilised *before* the silent agents died out.

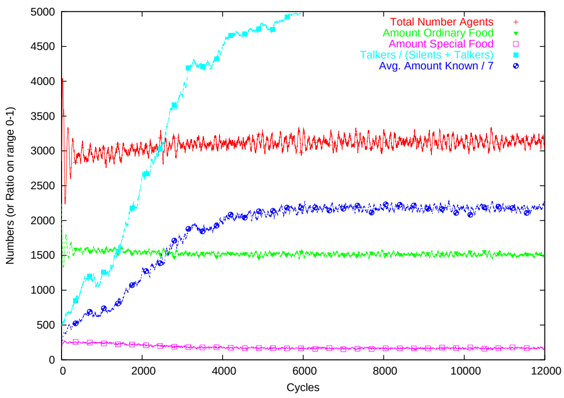


Figure 1: An example run where there are 6 special food types. Points are drawn approximately 10 generations apart.

Analysis shows there is no exclusive or common know-how. That is, for every one of the different things the agents can know, at any point in time, the percentage of agents that have this knowledge will fall in the same interval, yet there is never a point reach where all agents know all things (see for example Figure 1). Note that this means only that individual knowledge is distributed evenly at different points in time. Across particular regions of space some know-how may be shared by many agents while some could be held exclusively by a small number of individuals. In fact this is exactly what happens. We have equipped our model with a visual diagnostic tool to enable us to see which agents have the same know-how. Of course, having some type of know-how will lead to a higher energy level only if the agent stumbles upon the corresponding food. The speed of dissemination and the overall availability of knowledge depends on the density of communicating agents. Knowledge spreads faster in an environment that can support more agents. Also the numbers of communicating agents increases faster in an environment that has more special food (food that requires knowledge) in it.

5 Discussion and Future Work

5.1 The Effects of Communication

We have chosen a natural way of modelling the cost and benefit of communication. In our model the costs of communication are individual and temporally local. Also, the costs and benefits (for the receiving agents) are probabilistic. The hearing agent only benefits from the newl knowledge if it stumbles on a patch with the corresponding food. The cost for the communicating agent mirrors this. By compulsively giving out information the communicating agent enables its neighbours to take that same food. There may never actually be any competition if the food is not there. But when there is food in the vicinity, the neighbouring agents will be in the position to take it all. Nevertheless, compulsively giving free information is the evolutionarily successful strategy.

5.2 Selection

It may be argued that what we are showing here is kin selection. The agents live for a short time and they mostly take small steps. Under these circumstances an agent can be expected to have more contact with its offspring or parent than with any other particular random agents. These are simple agents which speak to every neighbouring agent the same, whether they are related or not, or even whether or not they are the same breed. Nevertheless, some biologists argue for kin rather than group selection even where the identity of kin may be 'mistaken', or in this case, not discriminated.

We can see no other explanation for why the silent agents die out except that, since they tend to be near their own relatives, they don't tend to get to know as much, and so they are out-competed for energy. Similarly, although a speaker gives up the advantage of its own knowledge in the short term, because it is likely to be in a community of similar agents, it is more likely to learn knowledge that helps it exploit more different kinds of food. What our simulations show is that, at least in some circumstances, this is the better strategy than free-riding.

There is obviously a great deal of work remaining to be done. It would be interesting to characterise more formally when and why the silents die out and the communicators dominate. We could also count how many of the communications going on are between relatives as a proportion of the whole. It would also be interesting to modify the model to explore things such as what happens if agents *do* restrict their communication to close relatives, practice deception, or just accidently, through their own ignorance, communicate useless or even harmful knowledge.

Nevertheless, the work as it stands brings interesting challenges to some existing Evolution of Language theory. For example, every theory on the emergence of language that presupposes that the cost to communication automatically means communication is not in itself adaptive needs to be reevaluated.

6 Conclusion

We have demonstrated that the propensity for communication can have a selective advantage despite being costly, provided that it has sufficient benefit to the community of speakers. The *act* of communicating may be costly to an agent, but the *propensity* to communicate will benefit the agent if it is consequently likely to learn from its children, parent, siblings and cousins. Further, we have shown that free riders who understand what they hear but do not share their own knowledge not only fail to inhibit the selection of free communicators, but will in the long term be out-competed if they tend to pass on their propensity for not sharing their knowledge and to live near their kin.

As we stated in our introduction, our results in no way challenge whether there are other selective forces that have affected the evolution of language, particularly language as we know it. But we have conclusively shown that arguments in favour of such mechanisms as selection for prestige cannot rely for evidence on free communication being non-adaptive. Communication may still have been the first selective advantage of language.

References

L. C. Aiello and P. Wheeler. The expensive-tissue hypothesis: The brain and the digestive-system in human and primate evolution. *Current Anthropology*, 36(2):199–221, April 1995.

Cristobal Baray. Effects of population size upon emergent group behavior. *Complexity International*, 6, 1998.

Lars-Erik Cederman and M. P. Rao. Exploring the dynamics of the democratic peace. *Journal of Conflict Resolution*, 45(6):818–833, 2001.

Daniel C. Dennett. *Kinds of minds: Towards an understanding of consciousness*. Weidenfeld and Nicolson, 1996.

Jean-Louis Dessalles. Language and hominid politics. In Chris Knight, Michael Studdert-Kennedy, and James Hurford, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 62–79. Cambridge University Press, Cambridge, UK, 2000.

Joshua M. Epstein and Robert Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press, MIT Press, Cambridge, MA, 1996.

Marc Esteva, Julian Padget, and Carles Sierra. Formalizing a language for institutions and norms. In Jean-Jules Meyer and Milinde Tambe, editors, *Intelligent Agents VIII*, volume 2333 of *Lecture Notes in Artificial Intelligence*, pages 348–366. Springer Verlag, 2001. ISBN 3-540-43858-0.

Charlotte K. Hemelrijk. Towards the integration of social dominance and spatial structure. *Animal Behaviour*, 59(5):1035–1048, 2000.

James Steele. What can archaeology contribute to solving the puzzle of language evolution?, April 2004. plenary talk at *The Evolution of Language*.

M.F. van Dartel, E. 0. Postma, and H.J. van den Herik. Universal properties of adaptive behaviour. In H. Blockeel and M. Denecker, editors, *Proceedings of the* 14^{th} *Belgium-Netherlands Conference on Artificial Intelligence (BNAIC '02)*, pages 59–66, Leuven, Belgium, 2002.

Uri Wilensky. The netlogo rabbits/grass/weeds model, 1998. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Uri Wilensky. *NetLogo*. Evanston, IL, USA, 1999.

A distributed view of language origins

Stephen J. Cowley
School of Psychology, University of Hertfordshire &
University of KwaZulu-Natal, South Africa
s.j.cowley@herts.ac.uk

Abstract

On a 'distributed' view, cognition is embodied, situated and develops as agents use the world to alter their cognitive powers. Among other things, this throws new light on how infants are transformed by talk. This process comes to be seen as one of mutual gearing where, acting jointly, adults and infants link affect-based behaviour with normative patterns. Next, conclusions are drawn for language- sensitive machines. Androids that orient to norms by using human dynamics will, it is suggested, throw new light on language origins. Apart from using powers based in a language faculty, they can be used to explore how the etiology of 'language' uses developmental processes that enable us to develop beliefs based on historical use of writing.

1 Introduction

"Life is .. but a tale told by an idiot, full of sound and fury, signifying nothing." Macbeth, 5.5. 17-25.

Writing with one eye to machines, I sketch how talk helps make babies into persons. Emphasis falls, not on the words that we find in speech, but rather the sound and fury of old tales. Adopting a distributed view of cognition (e.g. Hutchins, 1995; Clark, 1997), I follow Kirsh and Maglio (1994), in stressing how human intelligence uses real-time activity to alter the way cognition proceeds. From this perspective, the origins of language depend –not only on inner 'faculties' – but also how affect prompts infants to use visible and vocal expression.[1] While emphasising how caregiver *beliefs* prompt infants to change their multimodal human expression, I begin with the magical sound of speaking persons.

[1] The dynamics of utterance-activity use 'epistemic action' (Kirsh and Maglio, 1994). In the examples below a baby uses an adult to change its own neural states (sic).

2 Magical words?

Words seem magical because what we say is, often, only distantly related to how the words are taken. In conversation a person may say 'brown' or 'marrone' to mean roughly "How can that disgusting stuff possibly be chalk?" In spite of the gap between the said and what is meant, agreement is possible because, in real-time, agents gear to each other's expression (Cowley, 1997). In this paper, I sketch how the same abilities contribute to more primitive activity.

Infant response to a speaking person impacts on development by altering the cognitive processes that sustain human agency (Cowley, 2005). On this view, word-forms are second-order constructs rooted in kinetic, expressive and vocal events. In real-time, the infant responds affectively to persons whose dynamics are influenced by the norms applied in the circumstances. Using biomechanical biases, these cultural patterns allow infants to synchronise and anticipate events (Cowley, Moodley and Fiori-Cowley, 2004). Not only do they reflect on prior learning but, crucially, on caregivers' interpretations. Unknowingly, infants find that affect enables them to engage with the interaction order (Goffman, 1983; Drew and Wooton, 1988). As illustrated, dynamic joint behaviour or 'utterance-activity' connects affectively based behaviour with adult 'understanding'. In a world of cultural norms, while utterance-activity can be analysed as discourse, sentence-structure and words, its ontological origins lie in how bodies adjust to each other. Like most vertebrate communication, language is based in assessing and managing display (Owings and Morton, 1998). In development, as in evolution, changing assessments drive changes in management strategies.

3 Norms: the gearing hypothesis

Drawing on affective reactions, infants use biomechanics to prompt adults to act in line with folk psychological analyses. As they orient to the relevant norms, therefore, they depend on neither simple invariants nor pattern recognition. This case can be effectively presented around an example. While Distress Control Silence (DCS) routines are probably universal, they vary across cultures, relationships and interactions. They conform to a simple pattern:

- A baby shows distress
- A caregiver tries to control the baby
- The baby falls silent

Interestingly, as routines change, response is dominated by different aspects of caregiver regulation. While early DCS routines are based in kinesis (e.g. rocking, shaking and touching), work in three African communities showed that, at 14 weeks, this was already falling away in (some) isiZulu speaking dyads (Cowley et al., 2004). At this age, vocal and visible dynamics were often sufficient to prompt babies to do what the caregiver wanted. In other contexts, more use was made of picking the baby up, hugging and soothing. Social learning led some14 week olds to discover:

- Patterned vocalizations and gestures demand response.
- How and when to respond.
- What this affords.

Rather than appeal to what an infant 'knows', one can examine how each party regulates the other. On the infant side, in spite of sensory under-development and poor motor control, responses are managed by intrinsic motivation systems. Whatever their neural basis, the baby uses its brain for what Wilson (1998) calls 'prepared learning'.[2] Caregivers, by contrast, promote DCS routines because of how inner motives interact with cultural values, beliefs and strategies. In spite of this asymmetry, real-time sensitivity to affective display prompts infants to gear to adult expectations. Gradually, they develop regulatory strategies that allow each party to manipulate the other. Infants learn when micro-behaviour indexes likely affective reward (and punishment). Even before 4 months, babies are able to use utterance-activity to anticipate events. They can show sensitivity to *what another person wants*.

DCS routines have both selective value and major cognitive consequences. These arise as each party gears to the other. By scaffolding caregiver response, a baby alters the cognitive results of its actions. When falling silent on command, for example, an adult may find herself reacting to 'My baby is good', 'she really understands' or 'she is afraid of father'. DCS routines thus evoke normative behaviour. Especially when articulated, these give regularity and structure to a child's experience of the interaction order. In kwaZulu Natal, concepts like *thula* and *hlonipha* are powerful especially when, at DCS moments, a baby *fails* to fall silent. Because it does not thula (fall silent) a caregiver may treat it as failing to

[2] Cowley (forthcoming) appeals to Trevarthen's (1998; 2001) intrinsic motive formation. For discussion of empathy and perception-action mechanisms, see Preston and de Waal (2002).

show respect (hlonipha). The dyad's repertoire develops as interactions featuring co-ordination and conflict bind affect with expectations based on how, in different circumstances, adults use norms.

Development uses how a baby meshes its behaviour with adult beliefs and wants (e.g. be quiet if father is sleeping). Joint activity based in learning and norms shape the biomechanics of joint activity.

Co-ordination comes to reflect on locally favoured dynamic patterns that prompt culturally mediated joint action to promote human functionality. Much human behaviour is thus based in neither symbol processing nor pattern recognition but, rather, a kind of soft assembly. In this, external features –affective values associated with cultural procedures –recruit neural systems that subtend sensorimotor control. Although relying on statistical learning and biases (biomechanical and cultural), patterns emerge as infant motor control adjusts to normative dynamics. Social learning prompts dual control where, in some circumstances, infants incorporate adult beliefs into behavioural strategies (see, Spurrett and Cowley, 2004; Cowley, 2004b). In kwaZulu Natal, DCS routines help infants grasp dynamics based on expecting silence (thula) as a sign of respect (hlonipha).

4 Aligning to the words spoken

Since a baby uses manifest wants and beliefs, expressive behaviour enables cultural concepts to shape infant cognition. Even at 3 months, a child's behaviour is affected not just by utterance-activity but also important 'words'. That, however, is just a first step in aligning to what is said. Before considering this, however, it should be stressed that the baby is an attractive human body who manipulates its kin by using contingent behaviour and joint affect. The selective history that derives human functionality from utterance-activity is, therefore, likely to be ancient. That said, let us turn to another example.

The 9 month Luke is learning new ways of manipulating others. The baby, who knows no words, uses attentional abilities to pick up on signals based in how beliefs permeate his world. Consider other settings: were he a rural Kispigis speaker, he would be affectionately called a monkey and, perhaps, rely as much on manual as vocal communication. In Luke's belief-world, however, verbal intelligence is highly valued and his mother thinks (falsely) that a 'self' shows Luke what he can see, hear, feel etc. While mistaken, these culture specific beliefs shape many of her doings. Luke uses affect together with abilities to anticipate contingent patterning that prompts new ways of acting (see, Cowley, in prep). Below, he comes up with a proto-thought.

Fig 1. Luke's proto-thought ('get the block')

When Luke spontaneously generates a (feasible) novel form of behaviour, he self-rewards with a big smile. In thus altering his cognitive processes, he shows every sign of being close to becoming a self-regulating agent. While using time-locked co-ordination and joint routines, Luke does more than meet his mother's want. Though he does not know the source of his action, he grasps what she wants him *to do*. Whatever neural processes underlie the smile, they set off co-ordinated, self-initiated action. Immediately afterwards, he crawls off to get a block.[3]

How does Luke understand his mother? As English speaking adults, we note that, in 6 seconds, she *asks three times* if he wants to get the block. The words actually spoken, however, do not affect Luke. Lacking belief in words or meanings, as Cowley (in prep.) argues, he uses her body to decide 'what now?' Thus when analysed in real-time, mutual gearing can be shown to divide what happens into stages[4]

- Realising this is not a giving game.
- Grasping what it is all *about* (based on gaze following)
- Inhibiting other stimuli to promote joint activity
- Using her body to see what to do.

As at 3 months, events depend on close timing and inhibition of other stimuli. The baby can no also follow gaze, recognizes when joint activity is expected, crawl, and pick up desired objects. In 11 seconds, these schema come together and, anticipating his own doings, Luke rewards himself. This reward is a proto-thought or, in folk terms, "understanding that she wants him to fetch the block". In aligning to *what is actually said*, he effectively answers 'Do you want to fetch the block?"

Mutual gearing allows utterance-activity to transform human agency (Cowley, 2005). Soon, Luke will not only show understanding but also act in ways inviting analysis around words and rules (Spurrett and Cowley, 2004; Cowley, 2004b). As he babbles, he may make [fe] sounds that may, in some conditions, prompt his mother to 'fetching'. Given her beliefs, she may use this (falsely) as evidence that he 'knows' a word or meaning. Biomechanically, this affords new ways of integrating vocalizing with joint perception-action schema. Saying [fe] thus adds functionality to how he manipulates her. While based in associative learning, Luke's fetching differs from that of a dog. Affect laden signals enable him, unknowingly, to draw on beliefs and values. In learning to 'fetch', multimodal utterance-activity serves as a remarkable cognitive resource.[5]

5 The case for androids

Can a machine exploit social learning to discover word-magic? This empirical question can be addressed by harnessing computational powers to sensors and motor control systems. Given the power of mutual gearing, we can gain from rethinking the nature of language-sensitive machines. Infant behaviour shows that, among other things, these exploit behavioural dynamics jointly regulated by culture and affect. Each case of *falling silent*, *showing respect*, and *fetching that now* use circumstances and, for that reason, cannot be simulated by systems that are reliant on invariants or brute statistical learning. Indeed, it is because so much of human cognition is contingent that, to exploit utterance-activity, machines need to use micro-behaviour in sensing to norms.[6] If they do this, real-time human behaviour can be used to transform the machine's causal and cognitive powers.

Real-time sensitivity to multimodal utterance-activity will be possible if machines look like us and behave in human-like ways. Further, like infants, they will need to induce appropriate norm oriented human behaviour.[7] The engineering prob-

[3] This rolled away moments before. In the picture, his caregivers body 'points at the block'. To prompt him to get it she has been shifting her gaze from the block to Luke (and back again).

[4] In Tomasello's (1999) work, this kind of triadic behaviour is attributed to a special evolved socio-cognitive device (for critique, see Cowley, 2004a).

[5] There is no reason to think that Luke needs an inner process to represent forms isomorphic to those of "Do you want to fetch that?" In another sense, however, that *is* what he understands. His grasp of what to do uses how caregiver beliefs are *enacted*.

[6] It is instructive to contrast how humans and computers interact during Tetris (see, Kirsh and Maglio, 1994). While humans develop strategies, a computer runs a program. Obliterating this contrast, language-sensitive machines would use dynamics rather as do Tetris players.

[7] Empirically, machines will look and behave 'like' us. Rather than assume that they will be like babies, we should remember that laboratory trained bonobos learn elementary language. We

lem, therefore, is to design machines that use spontaneous micro-behaviour to build relationships. Only androids (MacDorman & Ishiguro, 2004) can induce behaviour whereby machines can use affect to alter their cognitive powers by using human wants and *de facto* beliefs (let alone explicit ones!) On the other hand, humans are often well-disposed to robots and, for social reasons, act to couple affective display with norm-based activity. Thus, in examining human-robot interaction, Cowley and Kanda (in press) comment, "Perhaps the most significant finding in how children respond to Robovie is that what they do is affected less by robot behaviour than an imagined relationship."

5 Magic in words

While babies do not understand the formal entities we associate with 'words', they are deeply affected by language-based beliefs. As Macbeth put it, they use the 'sound and fury' of human expression which, in other contexts, gives life to old tales. In one respect, then, the physical grounding of language lies in expressive patterns often regarded as 'idiotic'. How Luke uses his mother's body shows that dynamics function by altering *how cognition proceeds*. By gearing to his mother, who gears to him, he is able to produce a proto-thought that is followed up by activity that demands folk explanation (e.g. 'getting the block as was suggested').

For engineers, the distributed view has three main implications. First, instead of focusing exclusively on word-forms, they can explore how the interaction order uses bodily dynamics in affect-driven 'transparent' activity. Second, to use this spontaneously, we need machines that look like us, behave like us, and build human-like relationships. Third, development of androids may change our view of the evolution of language. Instead of invoking organs or instincts, the etiology of language seems to straddle nature, development and history.[8] Not only do its evolutionary roots allow babies to link world and brain-side development but, especially in literate communities, language comes to be institutionalised around word-use on linguistic beliefs that have altered in tandem with the historical processes that gave us technologies that include wax tablets, printed books and computer technology.

mutually gear to each other's strategic uses of affect (see, Cowley & Spurrett, 2003).

References

A. Clark, *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge MA, 1997.

S. Cowley, Of representations and language. *Language and Communication*, 17/4, 279-300, 1997.

S. Cowley, D. Spurrett, Putting apes, (body and language) together again. *Language Sciences*, 25: 289-318, 2003.

S. Cowley, Simulating others: the basis of human cognition? *Language Sciences,* 26/3: 273-299, 2004a.

S. Cowley, Contextualizing bodies: human infants and distributed cognition. *Language Sciences* 26, 565-591, 2004b.

S. Cowley, S. Moodley, and A. Fiori-Cowley, Grounding signs of culture: primary intersubjectivity in social semiosis. *Mind, Culture and Activity, 11/2:* 109-132, 2004.

[8] As Hauser et al. (2002) argue, 'narrow' aspects of language may draw on a selected neural systems for, say, recursion. In the distributed view, by contrast, emphasis is given to 'external language' or affect driven utterance-activity that that is affected by normative patterns. For discussion of the written bias of linguistics see Linell (1982); for a radical view of how writing has transformed language, see Harris (2000)

S. Cowley, Languaging: how babies and bonobos lock on to human modes of life. *International Journal of Computational Cognition, 3/1, 44-55,* 2005.

S. Cowley, Bridges to history: biomechanical constraints in language. To appear in N. Love (Ed.) *Integrational Linguistics and History,* Routledge, London, forthcoming.

S. Cowley, Brains and bodies in the world: connexions and word language. To appear in D. Moyal-Sharrock (Ed.) *Perspicuous Presentations*, in prep.

S. Cowley, T. Kanda, T. Friendly machines: Interaction-oriented robots today and tomorrow. To appear in *Alternation*, forthcoming.

P. Drew, A. Wooton, A. *Erving Goffman: Exploring the Interaction Order*. Polity Press, Cambridge, 1988.

E. Goffman, The Interaction Order*, American Sociological Review* 48: 1-17, 1983.

R. Harris, *Rethinking Writing*. Continuum, London, 2000.

M. Hauser, N. Chomsky, W. Fitch, W. The faculty of language: What is it, who has it and how did it evolve? *Science*, 298, 1569-1579, 2002.

E. Hutchins, *Cognition in the Wild*. MIT Press, Cambridge MA, 1995.

D. Kirsh, P. Maglio, On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 513-549, 1994.

P. Linell, *The Written Language Bias in Linguistics*. Linkoping University Press, Linkoping 1982

K. MacDorman, H. Ishiguro, The study of interaction through the development of androids. *IPSJ SIG Technical Reports 2004-CVIM-146,* 69-75. Tokyo, Japan, 2004.

D. Owings, E. Morton, *The Evolution of Vocal Communication: a New Approach*. Cambridge: Cambridge University Press, 1998.

S. Preston, F. de Waal, Empathy its proximate and ultimate bases. *Behavioral and Brain Sciences*, 25, 1-72, 2002.

D. Spurrett, S. Cowley, How to do things without words. *Language Sciences,* 26/5, 443- 466, 2004.

M. Tomasello, *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge MA, 1999.

C. Trevarthen, The concept and foundations of infant intersubjectivity. S. Braten (Ed.) *Intersubjective Communication in Early Ontogeny,* 15-46. Cambridge University Press, Cambridge, 1998.

C. Trevarthen, The neurobiology of early communication: Intersubjective regulation in human brain development. A. Kalverboer and A. Gramsbergen (Eds.) *Handbook of Brain and Behavior in Human Development,* 841-881. Kluver Academic, Dordrecht, 2001.

E. Wilson, *Consilience: The Unity of Knowledge,* Vintage Books, New York, 1998.

Examining Diversity in Cultural Learning

Dara Curran and Colm O'Riordan
*Computational Intelligence Research Group
IT Department
NUI,Galway
Ireland
dara.curran@nuigalway.ie
colmor@it.geminga.nuigalway.ie

Abstract

This paper examines the effect of cultural learning on a population of neural networks. We compare the genotypic diversity of populations employing only population learning and of populations using both population and cultural learning. We show that cultural learning is capable of achieving higher fitness levels and maintains a higher level of genotypic diversity.

1 Introduction

A number of learning models may be readily observed from nature and have been the focus of much study in artificial intelligence research. Population learning (i.e. learning which occurs at a population level through genetic material) is typically simulated using genetic algorithms. Life-time learning (i.e. learning which takes place during an organisms's life time through reactions with its environment) can be simulated in a variety of ways, typically employing neural networks or reinforcement learning models.

A relatively new field of study in artificial intelligence is synthetic ethology. The field is based on the premise that language and culture are too complex to be readily analysed in nature and that insight can be gained by simulating its emergence in populations of artificial organisms. While many studies have shown that lexical, syntactical and grammatical structures may spontaneously emerge from populations of artificial organisms, few discuss the impact such structures have on the relative fitness of individuals and of the entire population.

The focus of this paper is to attempt to understand the effect of cultural learning on a population of artificial organisms. This is accomplished by studying its effect on the population's fitness as well as its genotypic diversity. The remainder of this paper is arranged as follows. Section 2 introduces background research, including descriptions of diversity measures and cultural learning techniques that have been used in the past. Section 3 describes the experimental setup. Section 4 presents the Experiment Results and Section 5 presents conclusions.

2 Background research

2.1 Cultural Learning

Culture can be succinctly described as a process of information transfer within a population that occurs without the use of genetic material. Culture can take many forms such as language, signals or artifactual materials. Such information exchange occurs during the lifetime of individuals in a population and can greatly enhance the behaviour of such species. Because these exchanges occur during an individual's lifetime, cultural learning can be considered a subset of lifetime learning.

An approach known as synthetic ethology (MacLennan and Burghardt (1993); Steels (1997)) argues that the study of language is too difficult to perform in real world situations and that more meaningful results could be produced by modelling organisms and their environment in an artificial manner. Artificial intelligence systems can create tightly controlled environments where the behaviour of artificial organisms can be readily observed and modified. Using genetic algorithms, the evolutionary approach inspired by Darwinian evolution, and the computing capacity of neural networks, artificial intelligence researchers have been able to achieve very interesting results.

In particular, experiments conducted by Hutchins and Hazlehurst (Hutchins and Hazlehurst (1995)) simulate cultural evolution through the use of a hidden layer within an individual neural network in the population. This in effect, simulates the presence of a Language Acquisition Device (LAD), the physiological component of the brain necessary for language development, the existence of which was first suggested by Chomsky (Chomsky (1976)). The hidden layer acts as a verbal input/output layer and performs the task of feature extraction used to distinguish different physical inputs. It is responsible for both the perception and production of signals for the agent.

A number of approaches were considered for the implementation of cultural learning including fixed lexicons (Yanco and Stein (1993); Cangelosi and Parisi (1996)), indexed memory (Spector and Luke (1996)), cultural artifacts (Hutchins and Hazlehurst (1991); Cangelosi (1999)) and signal–situation tables (MacLennan and Burghardt (1993)). The approach chosen was the teacher/pupil scenario (Billard and Hayes (1997); Denaro and Parisi (1996); Cangelosi and Parisi (1996)) where a number of highly fit agents are selected from the population to act as teachers for the next generation of agents, labelled pupils. Pupils learn from teachers by observing the teacher's verbal output and attempting to mimic it using their own verbal apparatus. As a result of these interactions, a lexicon of symbols evolves to describe situations within the population's environment.

2.2 Diversity

Diversity measures typically quantify the differences between individuals in a population. It is commonly accepted that a population that is capable of maintaining diversity will avoid premature convergence and local maxima.

Diversity measures for populations of neural networks have been the focus of considerable research, focusing mainly on genotypic diversity (Y. Liu and Higuchi (2000); Opitz and Shavlik (1996); Brown (2003)). Many methods exist for the calculation of genotypic diversity, many based on binary representations. For the purposes of this research however, many schemes are unsuitable due to the nature of the marker-based encoding scheme used to represent each neural network.

Our scheme examines each block of the encoding and compares it to blocks of similar length in other encodings. Each encoding block contains a single node and a number of links emanating from that node. Since it would be difficult to compare arbitrary blocks of differing lengths to produce a diversity comparison, our approach pairs and compares blocks of similar lengths.

The diversity calculation follows the following steps: the genomes of the two individuals being compared are examined and blocks having similar lengths are identified and flagged in the genome of each individual. Each pair of blocks is then compared using a euclidian distance measure. The measure is then averaged over the length of the genome to produce a diversity measure.

The diversity calculation is applied such that every individual in the population is compared to all others, resulting in a global diversity measure for the entire population.

3 Experimental Setup

The following set of experiments each employs two populations. One population is allowed to evolve through population learning (by genetic algorithm), while the other employs both population and cultural learning. The experiments are carried out using a previously developed artificial life simulator (Curran and O'Riordan (2003)) capable of simulating population and cultural learning.

Cultural learning is implemented using a scheme developed by Hutchins and Hazlehurst (Hutchins and Hazlehurst (1991)) where the last hidden layer in a neural network functions as a verbal input/output layer. At the end of each generation, a number of the best individuals in the population is selected to instruct the next. Pupil networks observe teacher networks as they interact with their environment and at each stimuli, the teacher networks produce an utterance through their verbal I/O layer. The pupil responds to the utterance with its own, which is then corrected by back-propagation to more closely resemble the teacher's utterance. After the required number of these interactions (teaching cycles) have been completed, the teachers are removed from the population and the pupils continue to interact with their environment.

The problem domain for this set of experiments is the 5-bit parity problem. Each network is exposed to bit patterns and must determine whether the pattern represents an odd or even number. Fitness is assigned according to the mean square error of a network.

Each experiment consists of a population of 50 neural networks evolving for 250 generations with crossover and mutation rates set at 0.6 and 0.02 respectively. The population employing cultural learning takes the top 10% of each generation as teachers which interact with pupils for five teaching cycles. An additional parameter, cultural mutation, adds noise to each interaction with probability 0.02. Each of these parameters was determined empirically in previous work (Curran and O'Riordan (2004a,b)). The results presented are averaged from 20 independent runs.

4 Experiment Results

The graph presented in Figure 1 shows that, as shown in previous work, the addition of cultural learning is beneficial to the population. The average population error can be seen to be declining in both populations, but there is a clear distinction between the population employing cultural learning and the population using only population learning.

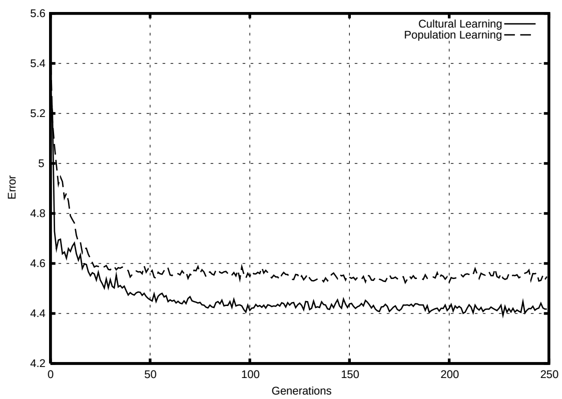


Figure 1: *Population Error*

Figures 2 and 3 show the average minimum and maximum error values for both populations. In the case of minimum error, the population employing cultural learning is generating individuals with lower error values than population learning alone.

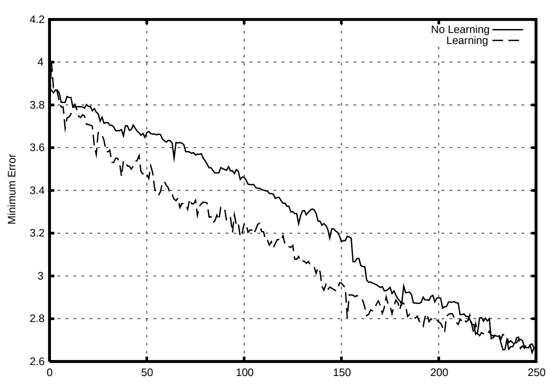


Figure 2: *Mininum Error Values*

Examining the maximum error graph it is clear that cultural learning is also generating individuals with a lower maximum error than population learning alone. Taking these two observations in combination, it can be said that cultural learning is not simply propping up the population (by improving the error values of weaker individuals), but that it is also capable of producing novel high performing individuals, thus improving the population's fitness as a whole.

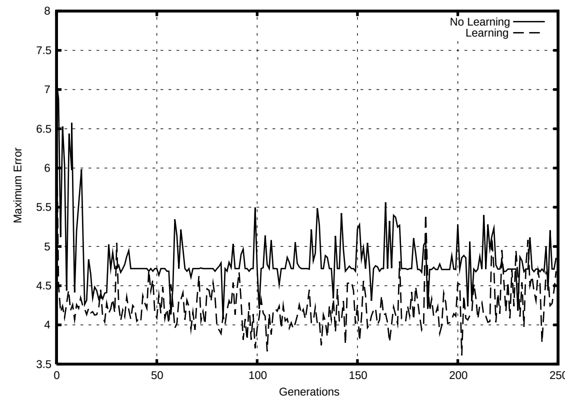


Figure 3: *Maximum Error Values*

The results for genotypic diversity, are illustrated in Figure 4. While both populations have a tendency to reduce diversity as the experiment progresses, the population employing only population learning is clearly converging faster. Cultural learning seems to be better at preserving genotypic diversity. In fact, it appears that genotypic diversity and fitness are positively correlated (0.82 value for Pearson's correlation) suggesting that a high level of diversity is desirable in the population.

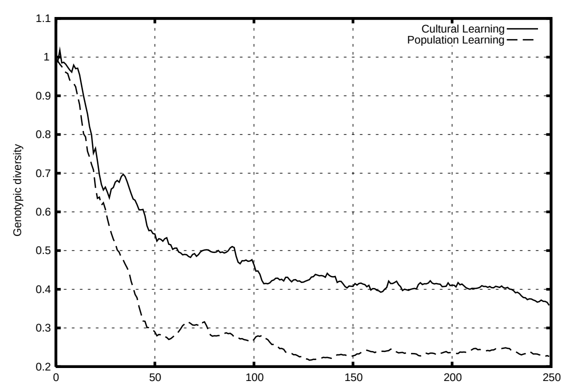


Figure 4: *Population Diversity*

To further understand the effect of cultural learning on the population, the population employing cultural learning was examined more closely. The average error value was taken at the start of each generation (before any teaching had taken place) and compared to the error value obtained subsequent to teaching. The results presented in Figure 5 are especially interesting when further compared to the error value obtained by the population using only population learning.

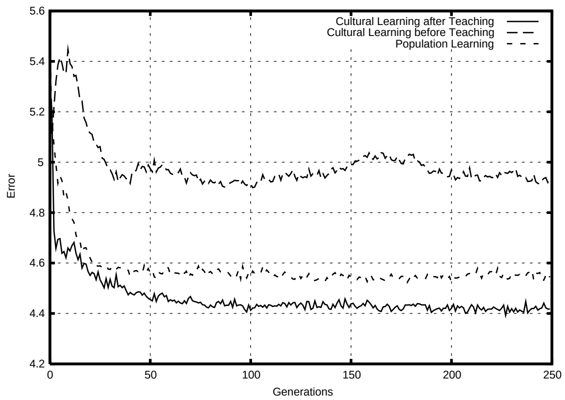


Figure 5: *Hiding Effect*

It appears that before individuals have the opportunity to learn through teaching, they perform very poorly compared to their population learning counterparts. However, once teaching is applied, they surpass the performance of population learning. This masking of what are infact mediocre,or poor individuals by another process has been previously identified as the Hiding Effect (Mayley (1997)).

Since its definition, the hiding effect has been frequently considered the anti-thesis of the Baldwin effect (Baldwin (1896)) where learning guides the process of evolution. However, these results suggest that there may be another aspect of the hiding effect: by masking poor genetic material, the population is effectively protecting its diversity and avoiding premature convergence.

5 Conclusions

It is clear from these results that cultural learning is having a direct and significant impact on the performance of the population for this particular problem set, namely reducing the average error of individuals in the population. In addition, from examination of maximum and minimum error values within each population it can be said that cultural learning achieves its higher fitness performance not only through the correction of weak individuals, but through a process of novel individual creation.

We have shown through these experiments that there are new aspects to the hiding effect which may not have been previously detected. By preserving initially mediocre individuals, the population is capable of maintaining its diversity and boost its performance. The genetically mediocre individuals are then allowed to learn through cultural learning, attaining fitness levels higher than what would be possible using population learning alone.

In future work we will address more complex problems and examine how the addition of cultural learning may aid where environments are changing dynamically.

Acknowledgements

This research is funded by the Irish Research Council for Science, Engineering and Technology.

References

J.M. Baldwin. A new factor in evolution. In *American Naturalist 30*, pages 441–451, 1896.

A. Billard and G. Hayes. Learning to communicate through imitation in autonomous robots. In *7th International Conference on Artificial Neural Networks*, pages 763–738, 1997.

G. Brown. *Diversity in Neural Network Ensembles*. PhD thesis, University of Birmingham, 2003.

A. Cangelosi. Evolution of communication using combination of grounded symbols in populations of neural networks. In *Proceedings of IJCNN99 International Joint Conference on Neural Networks (vol. 6)*, pages 4365–4368, Washington, DC, 1999.

A. Cangelosi and D. Parisi. The emergence of a language in an evolving population of neural networks. *Technical Report NSAL–96004, National Research Council, Rome*, 1996.

N. Chomsky. On the nature of language. In *Origins and evolution of language and speech*, pages 46–57. Annals of the New York Academy of Science, New York. Vol 280, 1976.

D. Curran and C. O'Riordan. On the design of an artificial life simulator. In V. Palade, R. J. Howlett, and L. C. Jain, editors, *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES 2003)*, University of Oxford, United Kingdom, 2003.

D. Curran and C. O'Riordan. A comparison of population learning and cultural learning in artificial life societies. In *ALIFE9*, Boston, Massachusetts, 2004a.

D. Curran and C. O'Riordan. Cultural evolution for sequential decision tasks: Evolving tic-tac-toe players. In *GECCO 2004*, Seattle, Washington, 2004b.

D. Denaro and D. Parisi. Cultural evolution in a population of neural networks. In *M.Marinaro and R.Tagliaferri (eds), Neural Nets Wirn-96.New York: Springer*, pages 100–111, 1996.

E. Hutchins and B. Hazlehurst. Learning in the cultural process. In *Artificial Life II, ed. C. Langton et al.* MIT Press, 1991.

E. Hutchins and B. Hazlehurst. How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte, editors, *Artificial Societies: The Computer Simulation of Social Life*, pages 157–189. UCL Press: London, 1995.

B. MacLennan and G. Burghardt. Synthetic ethology and the evolution of cooperative communication. In *Adaptive Behavior 2(2)*, pages 161–188, 1993.

G. Mayley. Guiding or hiding: Explorations into the effects of learning on the rate of evolution. In *Proceedings of the Fourth European Conference on Artificial Life*. MIT Press, 1997.

D. W. Opitz and J. W. Shavlik. Generating accurate and diverse members of a neural-network ensemble. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 535–541. The MIT Press, 1996.

L. Spector and S. Luke. Culture enhances the evolvability of cognition. In *Cognitive Science (CogSci) 1996 Conference Proceedings*, 1996.

L. Steels. The synthetic modeling of language origins. In *Evolution of Communication*, pages 1–34, 1997.

X. Yao Y. Liu and T. Higuchi. Evolutionary ensembles with negative correlation learning. In *IEEE Transactions on Evolutionary Computation*, volume 4(4), pages 380–387, 2000.

H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots. In *From Animals to Animats 2. Proceedings of the second International Conference on Simulation of Adaptive Behavior*, pages 478–485. MIT Press, Cambridge Ma., 1993.

The co-evolution of language and music

W Tecumseh Fitch
School of Psychology, University of St. Andrews
St. Andrews, Scotland
wtsf@st-andrews.ac.uk

Abstract

After a long hiatus, the evolution of language has become a major topic of research and discussion, with considerable empirical and theoretical progress made in the last decade. The same is less true of the evolution of music, despite a recent surge of interest in this topic. Like language, music is a human universal, found in all cultures, which has phrase structure and entails learning and cultural transmission. In contrast, music serves no obvious functional purpose. According to Darwin, the human musical faculty "must be ranked amongst the most mysterious with which he is endowed". This makes the evolution of music an interesting puzzle for evolutionary biologists.

In this talk I discuss the possibility that the deep similarities between language and music indicate a shared evolutionary history. In particular, the fact that both language and music are human universals, have, suggests that any theory of the evolution of language will have implications for the evolution of music, and vice versa. I first discuss the similarities and differences between language and music, focusing on mechanisms of music perception and the ontogeny of prosodic communication, and discussing comparative data regarding various animal communication systems commonly called musical (such as bird and whale "song"). After briefly discussing possible functions of human music (courtship, group cohesion, mother/infant communication) I will end by discussing the phylogenetic history of music. I conclude that many strands of evidence support Darwin's (1871) hypothesis of an intermediate stage of human evolutionary history. This hypothetical stage was characterized by a communication system that resembled music more closely than language, but was identical to neither. This pre-linguistic system, which I call "prosodic protolanguage", provided a precursor for both modern language and music.

This hypothesis leads to some empirically testable predictions. In particular, the considerable inter-individual variation in human musical skill makes it an excellent system to study genetic and neural mechanisms (comparing musicians to non-musicians) underlying a complex cognitive universal of our species.

The Emergence of MetaCommunicative Interaction: some theory, some practice

Jonathan Ginzburg[⋆]
[⋆]Department of Computer Science
King's College, London
The Strand, London WC2R 2LS
ginzburg@dcs.kcl.ac.uk

Zoran Macura[†]
[†]Department of Computer Science
King's College, London
The Strand, London WC2R 2LS
macura@dcs.kcl.ac.uk

Abstract

A central concern of work on the evolution of language has been to offer an account for the emergence of syntactically complex structure, which underwrites a compositional semantics. In this paper we consider the emergence of one class of utterances which illustrate that semantic expressiveness is not correlated with syntactic complexity, namely metacommunicative interaction (MCI). These are utterance acts in which conversationalists acknowledge understanding or request clarification. We offer a simple characterization of the incremental change required for MCI to emerge from an MCI-less linguistic interaction system. We discuss the evolutionary background in which MCI might arise and become adaptive. We describe a system in which computational simulations can be run designed to test hypotheses regarding the emergence of MCI.

1 Introduction

A central concern of work on the evolution of language has been to offer an account for the emergence of syntactically complex structure, which underwrites a compositional semantics. In natural language, semantic expressiveness is not correlated with syntactic complexity. A key feature of natural language, which provides striking instance of *syntactically underdetermined* semantic complexity, is metacommunicative interaction (MCI)—utterance acts in which conversationalists acknowledge understanding or request clarification. (1b) exemplifies such a syntactically simple form which, nonetheless, in context can acquire a highly complex content:

(1) (a) A: Did Bo leave?

(b) B: Bo?; ("Bo?" can mean in this context *Are you asking if Bo, of all people, left* or *who were you referring to as Bo?*).

Indeed NL possess forms whose sole meaning concerns MCI, as exemplified by (2), a form whose sole use is to query an antecedently uttered polar interrogative whose subject has unclear reference:

(2) Do I like who?

The need to verify that mutual understanding among interlocutors has been achieved with respect to any given utterance—and engage in discussion of a clarification request if this is not the case—is one of the central organizing principles of conversation (Schegloff (1992); Clark (1996)). However, hitherto there has been little work on the emergence of MCI meaning. Communicative interaction is fundamental to EoG(rammar) work, since it is interactions among communicating agents that leads an initial 'agrammatical' system to evolve into a grammar (with possible, concomitant phylogenetic modification; see e.g. Briscoe (2000); Kirby (2000)). However, given an I-language perspective, the communicative aspect as such is not internalized in the grammar (though see Steels (1998)). Consequently, such models of EoG cannot explain the existence of forms whose meaning is intrinsically MCI oriented.

In this paper we offer a simple characterization of the incremental change required for MCI to emerge from an MCI-less CIS. We discuss the evolutionary background in which MCI might arise and become adaptive. Finally, we report on a system we are developing in which computational simulations can be run designed to test hypotheses regarding the emergence of MCI.

2 Metacommunicative Interaction and EoL

2.1 The significance of MCI for a linguistic community

By metacommunicative interaction one means any interaction that comments about the communicative process underlying an utterance. More specifically, the commonest MCI utterances are: acknowledgements that an utterance has been understood, clarification requests (CRs) in which an unclear aspect of the utterance is queried, and corrections, where indications are provided of erroneous assumptions concerning naming, concepts associated with predicates etc. (3), from the London Lund corpus, contains a CR (utterance (2)), a correction (utterance (4)), and an acknowledgement (utterance (5)):

(3) A(1): did you also scotch that other story which is something like was he wasn‘t he refused the chair in Oxford

a(2): who

A(3): Skeat, wasn‘t he refused

a(4): that‘s Meak

A(5): oh Meak, yes

(London Lund S.1.9, p. 245)

What significance does MCI have for linguistic interaction within a community? MCI is redundant in so far as the communication channel, i.e. that which mediates between speaker and addressee, is perfect or close to that. The need for MCI arises when the communication channel is intrinsically liable to breakdown. If NL resembled formal languages like first order predicate calculus (as often implicitly assumed in EoL work, see e.g. Kirby (2000)), then problems with the communication channel would be restricted to actual physical problems with the speech signal (mishearing, mispronunciation, noise and the like), problems that affect just about any naturally occurring communicative interaction system. However, NL diverges radically from first order predicate calculus in its context dependence. This manifests itself in (at least) three phenomena:

(4) a. **indexicality**: words like ‘I’, ‘You’, ‘here’, ‘now’, that are resolved relative to the ongoing speech situation.

b. **anaphoricity**: words and phrases that are resolved relative to semantic values established by previous utterances (e.g. pronouns, non sentential utterances etc).

c. **ambiguity**: words and phrases which possess multiple senses, one of which is utilized in a given context.

Moreover, even a language like first order predicate calculus used by agents who can reflect about intentions underlying communicative acts, will give rise to the sort of inferences that have come to be known as Gricean conversational implicatures (Grice (1989)). These add an extra layer of uncertainty to the communicative process.

Given this, acknowledgements, CRs and corrections are a key communicative component for a linguistic community. They serve as devices for allaying worries about miscommunication (acknowledgements) or for reducing mismatches about the linguistic system among agents (CRs and corrections). That is, they serve as a device for ensuring a certain state of equilibrium or lack of divergence gets maintained within a linguistic community.

2.2 The Emergence of MCI: basic ingredients

Given the importance that MCI has for linguistic interaction, some fundamental questions that need to be answered are:

(5) a. Under what circumstances does a linguistic interaction system without MCI evolve into one that has MCI?

b. What mechanisms are involved in such a development?

c. Why is the resulting interaction system maintained?

In order to address these issues, we need to fix what we mean by *an interaction system with MCI*. In the literature on the semantics and pragmatics of dialogue—more on which see section 3— a number of interaction systems have been defined where in addition to the regular illocutionary acts (assertion, querying, commanding etc), also additional grounding acts (e.g. acknowledgements) are available (see e.g. Poesio and Traum (1997)) and also systems where clarification requests are available (see e.g. Ginzburg and Sag (2000); Ginzburg and Cooper (2004)). Such systems assume that as a preliminary to the processing of an utterance u an addressee A checks whether she understands u. If she does, A optionally responds with an acknowledgement, and then reacts in the conventional way to the utterance (accepting/disputing an assertion, answering a query, and so on.). On the other hand, if A does not fully understand u, A poses a query that requests clarification concerning the unclear aspect of u (e.g. inability to resolve a referent, unfamiliarity with or mishearing of a word, etc.) using a number of predefined operations on utterances and utterance meanings.

Poesio and Traum (1997); Ginzburg and Sag (2000); Ginzburg and Cooper (2004) show how existing formal frameworks for grammatical/semantic processing of MCI-less natural language can be extended to process natural language that includes MCI utterances such as acknowledgements and CRs. To understand what is involved, though, one can restrict attention to much simpler systems. We mention two here discussed in Ginzburg (2001).

The utt(erance) ack(nowledgement) game In this game, given an utterance u_0 consisting of a string (word$_1 \ldots$,word$_i$,$\ldots$,word$_n$) by the master, the novice may respond with the utterance u_1: $word_i$. In this context, this utterance is assigned content: *novice acknowledges that an utterance including the word word$_i$ happened*. This fact now becomes part of the novice's and master's common ground. What capabilities does playing *utt-ack* game require from the novice?

- Phonological imitation and segmentation module (can be played in one word mode, i.e. game does not require novice to have syntactically complex capabilities).
- Ability to form mutual beliefs.

The reward for playing this game is shared interaction with the master. Who can play this game?

- Human neonates: the initial stage of speech consists largely in playing this game. Bates (1979); Ninio and Snow (1996)
- Chimps: Greenfield and Rumbaugh (1993).

A rudimentary game with CRs: The ack-huh? game Given an utterance u_0, the responder may acknowledge the utterance or pose a simple CR querying the content of u_0? For instance:

(6) Master: You want the ball? Novice: (i)huh?/(ii)ball?

What additional capabilities does playing *ack-huh* require from the novice?

- Querying
- the ability to form questions querying the contents of antecedently uttered utterances.
- No requirement for syntax

Who can play this game?

- Human neonates (from approx 20 months)
- **Not** chimps: Greenfield and Rumbaugh (1993)

The key feature of these games is at the level of ontology, namely the possibility of reference to utterances and sub-utterances and their properties. In particular, for the *ack-huh?* game agents require a notion of synonymy between utterances (i.e. the ability to reformulate in a way that preserves content), otherwise any metacommunicative-oriented discussion will be circular. Thus, the simplest agent with the ability to *discuss* a CR is an agent who can communicate contents such as "I don't understand (previous-utterance)" and "What do you mean (previous-utterance)". Given an agent who can reflect and form questions about entities in the domain, this means that once 'say' and 'mean' predicates are in the language, then basic clarification requests can be expressed. Consequently, the emergence of metacommunicative interaction-oriented utterances that go beyond mere acknowledgement, as exemplified in the *ack-huh?* game, can be viewed as an instance of the problem of how vocabulary emerges to talk about a class of entities in a domain, given the need/desire to do so.

However, as should be obvious from this discussion, metacommunicative interaction *slows down* interaction—it involves expending effort away from the actual topic of discussion (food, social relations, danger etc) onto mere conversation management. Hence, it is *prima facie* clear that it is not adaptive in a small, tight knit community, where lexical differences among adults are minimal and where it is extremely unlikely that issues of reference resolution will arise. This *suggests* that metacommunicative interaction could only emerge in a linguistic community where members have significantly diverse experiences, which could lead to lexical divergences arising (e.g. nomads who share territory on a seasonal basis). We speculate that MCI has emerged as an interactional device that keeps members of a linguistic community from diverging too widely from each other's linguistic capabilities, say in terms of their basic vocabulary.

The plausibility of this speculation can be assessed by converting it into more concrete questions such as the following:

(7) (a) In a community with minor but random lexical differences where some people use clarification requests, whereas others do not, do the clarification request users gain an advantage?

(b) Given a community A where clarification requests do not get expressed, and community B where they do, how do the two communities evolve with respect to vocabulary drift?

In the following sections we discuss a computational system we have been developing. This system is intended to run simulations that can address questions such as (7). Experimental results from these simulations will be presented at the workshop.

3 Modelling dialogue

3.1 Issue-based Dialogue Management

In this section we outline some of the basic principles of Issue-based Dialogue Management, which underpins our dialogue system.

Information States We assume information states of the kind developed in the KoS framework (e.g. Ginzburg (1996); Larsson (2002)) and implemented in systems such as GODIS and CLARIE (see e.g. Larsson (2002); Purver (2004)). On this view each dialogue participant's view of the common ground is structured by a number of attributes including the following three: FACTS: a set of facts representing the shared assumptions of the CPs, LATESTMOVE: the most recent grounded move, and QUD ('questions under discussion'): a set—often taken to be structured as a stack—consisting of the currently discussable questions.

Querying and Assertion Both querying and assertion involve a question becoming maximal in the querier/asserter's QUD: the posed question q for a query where q is posed, the polar question $p?$ for an assertion where p is asserted. Roughly, the responder can subsequently either choose to start a discussion (of q or $p?$) or, in the case of assertion, to update her FACTS structure with p. A dialogue participant can downdate $q/p?$ from QUD when, as far as her (not necessarily public) goals dictate, sufficient information has been accumulated in FACTS. The querying/assertion protocols (in their most basic form) are summarized as follows:

(8) cooperative query exchange

1. `LatestMove.Cont = Ask(A,q): IllocProp`
2. `A: q becomes QUD maximal; release turn`
3. `B: q becomes QUD maximal; take turn; make q-specific utterance;`[1] `release turn.`

(9) cooperative assertion exchange

1. `LatestMove.Cont = Assert(A,p): IllocProp`
2. `A: p? becomes QUD maximal, release turn`
3. `B: p? becomes QUD maximal, take turn;` $\langle$ `Option 1: Discuss p?, Option 2: Accept p` $\rangle$

(10)
1. `LatestMove.Cont = Accept(B,p) : IllocProp`
2. `B: increment FACTS with p; pop p? from QUD;`
3. `A: increment FACTS with p; pop p? from QUD;`

Grounding Interaction Grounding an utterance $u : T$ ('the sign associated with u is of type T') is modelled as involving the following interaction. (a) Addressee B tries to anchor the contextual parameters of T. If successful, B acknowledges u (directly, gesturally or implicitly) and responds to the content of u. (b) If unsuccessful, B poses a Clarification Request (CR), that arises via *utterance coercion* (see Ginzburg and Cooper (2001)). For reasons of space we do not formulate an explicit protocol here—the structure of such a protocol resembles the assertion protocol.

3.2 CLARIE

CLARIE (see Purver (2004)) is a dialogue system using the TRINDIKIT framework and is based on the KOS approach to dialogue modelling.[2] It is designed to be able to (a) interpret users' clarification questions and respond suitably, and (b) ask clarification questions in order to learn new words and phrases. It is based on the dialogue system GoDiS, but uses

[1] An utterance whose content is either an answer to q/or a question q_1 which is a subquestion of q.

[2] TRINDIKIT Larsson and Traum (2000) is a toolkit for building and experimenting with dialogue move engines and information states. TRINDIKIT allows a modular structure where various models can feed information into the IS, or can read information from it. The TRINDIKIT also allows *resources* to be defined, modules which don't play a particular part in the overall control algorithm but which provide particular information or capabilities which can be called upon the other modules.

a Head Driven Phrase Structure grammar (Ginzburg and Sag (2000)) for interpretation and generation. It also incorporates the ellipsis resolution techniques of SHARDS (see Fernández et al. (forthcoming)).

4 An ALife Simulation

We are currently running artificial life simulations on a population of agents with dialogue capacities based on CLARIE. The simulation part is built up using RePast (developed by ROAD), a set of Java libraries that allow programmers to build simulation environments. The running of the simulation is divided into time steps or 'ticks', and at each tick some action occurs using the results of previous actions as its basis. Agents are created and placed in an environment in which they are able to wander around in search of food resources. Agents are endowed with a vision capacity in order to see food resources as well as other agents. Upon meeting another agent, the two agents enter a dialogue.

In order to test the questions raised in (7) the agents need to have minor but random lexical differences, and clarification requesting (CR) capabilities. As a working assumption we build in variation along the agents' lexicons restricted to nouns. Some motivation for this comes from the fact that verb clarification is very rare, where the CR ratio is 40:1 for nouns vs. verbs (Purver 2004).

Two types of agents exist in the model; agents capable of making a clarification request and those incapable of doing so. When an utterance is passed from one agent to another, the agent receiving the utterance tries to parse it using its grammar (which is same for every agent). If the utterance contains an unknown word its semantic interpretation fails in the grounding process. This triggers one of the two events (depending on the agent's clarification capability): a clarification request (CR) is produced, or a lexicon acquisition algorithm (LAA) is called. If a clarification question is produced, the answer to that question is used to resolve the semantics of the unknown word. If the lexicon acquisition algorithm is called the agent tries to ground the unknown word without producing a clarification request.

In the simulations agents are distributed randomly in the environment at the start of each simulation run. Each agent is associated with its own version of CLARIE. Once the simulation starts the agents begin walking randomly in the environment. Agents are able to perceive other agents that fall within their field of vision. Once two agents see each other, they engage in a dialogue by calling their dialogue systems respectively. After the completion of the dialogue the agents continue walking.

5 Conclusions and Future Work

In this paper we have discussed how metacommunicative interaction (MCI) serves as a key component in the maintenance of a linguistic interaction system. We have outlined the basic components that need to emerge in order that an MCI-less linguistic system evolves into an MCI-containing system. We have described the system we are developing for experimentally assessing the emergence scenario of MCI.

We believe that one of the most significant pay offs that studying the emergence of MCI will give rise to is the need to tackle the issue of the costs and rewards of participating in dialogue. Some initial steps in this direction have been taken by Young (2002); Paek and Horvitz (1999).

Acknowledgements

This work is supported by an EPSRC quota award to the second author and by grant RES-000-23-0065 from the Economic and Social Council of the United Kingdom to the first author. We would like to thank two reviewers for EELC05 for their very helpful comments.

References

Elizabeth Bates. *The Emergence of Symbols*. Academic Press, 1979.

E. J. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.

Herbert Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

Raquel Fernández, Jonathan Ginzburg, Howard Gregory, and Shalom Lappin. SHARDS: Fragment resolution in dialogue. In H. Bunt and R. Muskens, editors, *Computing Meaning*, volume 3. Kluwer, forthcoming.

Jonathan Ginzburg. Interrogatives: Questions, facts, and dialogue. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, Oxford, 1996.

Jonathan Ginzburg. The emergence of metacommunicative meaning, 2001. Paper presented at the 13th Amsterdam Colloquium on Semantics and Logic, December 2001.

Jonathan Ginzburg and Robin Cooper. Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the Association for Computational Linguistics*, Toulouse, 2001.

Jonathan Ginzburg and Robin Cooper. Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, 27(3):297–366, 2004.

Jonathan Ginzburg and Ivan A. Sag. *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California, 2000.

Patricia Greenfield and Sue Savage Rumbaugh. Comparing communicative competence in child and chimp: The pragmatics of repetition. *Journal of Child Language*, 20:1–26, 1993.

Herbert Paul Grice. *Studies in the Ways of Words*. Harvard University Press, 1989. Reprinted from a 1957 article.

S. Kirby. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, 2000.

S. Larsson and D. Traum. Information state and dialogue management in the trindi dialogue move engine toolkit. In *Natural Language Engineering Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*, pages 323–340. Cambridge University Press, 2000.

Staffan Larsson. *Issue based Dialogue Management*. PhD thesis, Gothenburg University, 2002.

Anat Ninio and Catherine Snow. *Pragmatic Development*. Westview, 1996.

T. Paek and E. Horvitz. Uncertainty, utility, and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In *AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, Cape Cod, MA., November 5-7 1999.

Massimo Poesio and David Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3), 1997.

Matthew Purver. *The Theory and Use of Clarification in Dialogue*. PhD thesis, King's College, London, 2004.

ROAD. Recursive porus agent simulation toolkit. http://repast.sourceforge.net

Emanuel Schegloff. Repair after the next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5):1295–1345, 1992.

L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1(2):169–194, 1998.

S. Young. Talking to machines (statistically speaking). In *Int Conf Spoken Language Processing*, Denver, Colorado, 2002.

Learning and Transition of Symbols in Chaotic Dynamical System: Toward Dynamical Model of Symbolic Individual

Takashi Hashimoto⋆
⋆ School of Knowledge Science
Japan Advanced Institute of
Science and Technology (JAIST)
1-1, Tatsunokuchi, Ishikawa,
Japan, 923-1292
hash@jaist.ac.jp

Akira Masumi†
† School of Knowledge Science
Japan Advanced Institute of
Science and Technology (JAIST)
1-1, Tatsunokuchi, Ishikawa,
Japan, 923-1292
a-masumi@jaist.ac.jp

Abstract

In this paper, we aim at constructing a model for agent showing symbolic activities. Based on a discussion of a concept of symbol systems from the viewpoint of dynamical systems, we propose correspondences of symbolic activities to behaviour of dynamical systems, such as, symbols to attractors, symbol manipulations to transitions among attractors, and manipulation rules to order of the transition. Thus, by using a dynamical system showing transition among attractors, we may be able to represent symbolic activities, in part. We try to construct such a system with a chaotic neural network that is a coupled NZ maps. We confirm that this model can memorize several patterns as attractors of the system and shows transition among the memorized patterns. Further, we exemplify that not only recall of the memorized patterns but also transitions among them are induced by certain external inputs. Examining these simulation results supports that the coupled NZ maps are used as a model of symbolic agent for the study of emergence of linguistic communication.

1 Introduction

The remarkable feature of linguistic communications is to use symbols for transmitting information and mutual understanding. Deacon (1997) pointed out that humans are symbolic species, namely, we show symbolic cognitive activities such as learning, formation, and manipulation of symbols. In the research of the origin and the evolution of language, we should elucidate the emerging process of such symbolic cognitive activities.

Most of agent models in simulation studies of the language evolution presuppose the symbol processing ability (Cangelosi and Parisi, 2002). For example, a computational model for the evolution of compositional syntax introduced in (Kirby and Hurford, 2002; Kirby, 2002) can possess grammatical rules representing the correspondence of meanings and character strings that are considered as combinations of symbols. In a dynamical systems model for the evolution of prototypical category structure introduced in (Hashimoto, 2002a), agents can emit, receive and process sequences of words.

In order to understand the origin of language, we should deal with the emerging process of such ability of symbol processing. To this end, we need a model of agent that autonomously acquires the ability of symbolic cognitive activities for effectively studying the emergence and the evolution of linguistic communication with the constructive approach.

The mathematical and computational studies of symbol processing have been done in artificial intelligence and connectionism. The former has difficulty in the self-organization and emergence of symbols, since symbols and the syntactic rules governing the symbol processing are usually given by hand. While the latter can acquire, in part, the symbolic representation from scratch without preparing explicit symbolic elements, it is not good at explicitly describing symbols and their processing rules, since symbols are in nature distributedly represented in neural networks. Thus, a new approach for symbol formation has been required (Harnad, 1990). A new approach is often supposed to be an integration of both artificial intelligence type and connectionism type methods[1].

A recent development to view cognitive systems is the dynamic perspective. van Gelder (1998) argued

[1]Note that using connectionism model does not necessarily mean that no symbolic element is prepared.

that cognitive systems can be understood well by considering them as dynamical systems. This viewpoint is also progressed to describe dynamic aspects of brain activities using the framework of dynamical systems and chaos (Tsuda, 2001).

The purpose of this research is to construct a model of agent with a dynamical system, which shows symbolic cognitive activities. When we construct a model of cognitive agent for linguistic communication, we take the dynamic viewpoint not only for the cognitive systems (van Gelder, 1998) but also for language (Hashimoto, 2002a,b). The dynamic view for language means that symbols are not mere correspondence of words to referents and symbol formation is not mere assigning process of words to some objects.

The rest of this paper is organized as follows. In §2 we discuss how symbolic systems are able to be represented in dynamical terms. Based on the discussion, we propose a model of dynamical systems for symbol formation in §3. Concretely, the model is a coupled chaotic dynamical systems. The simulation results of the model are shown in §4. We discuss the results in §5 and conclude this paper in §6.

2 Symbol Systems as Dynamical Systems

To render the symbolic activities in the framework of dynamical systems, we consider features of symbols. In general, symbols are considered to represent or to signify something and are manipulated according to some rules such as a grammar in languages or a deduction rule in calculation and formal thought.

Harnad (1990) summarizes the feature of symbol systems as the following definition:

1. A symbols is a set of arbitrary physical tokens that are
2. manipulated on the basis of explicit rules
3. that are likewise physical tokens and strings tokens.
4. The rule-governed symbol-token manipulation is based purely on the shape of the symbol tokens and
5. consists of rulefully combining and recombining symbol tokens.
6. There are primitive atomic symbol tokens and
7. composite symbol-token strings.
8. The entire system and all its parts are all semantically interpretable: The syntax can be systematically assigned a meaning.

This definition describes a system that can be interpreted as symbolic rather than an internal symbolic activities. In order to construct an agent model that shows symbolic activities, we construe this definition as internal cognitive processes. Further, to implement the agent model using a dynamical system, we interpret the processes with the concepts of dynamical systems.

The items 1, 6 and 8 say that there are some entities that are accepted or interpreted as representing something such as objects, states of affairs, or abstract ideas by cognitive agents. A cognitive process to receive some physical tokens evokes a cognitive process to recall some memorized concepts. In the terms of dynamical systems, this process can be viewed as that some inputs to a dynamical system bring the system to certain states. This representative function is thought of as being realized by a kind of memory that is usually modeled as attractors of the dynamical system.

The items 2, 3 and 4 imply that the cognitive agent performs symbol manipulation as a process of successive recall of concepts (memories) and the successions are rule-governed. In the dynamical terms, we conceive the process as (spontaneous) transitions among attractors in which the transition is rule-governed or, at least, ordered.

The items 5, 7 and 8 mean that a part of sets or some series of physical tokens, not all sets and series, are accepted as a ordered combination of entities, not as independent entities, and receiving processes of such series induce retrieval processes of concepts. These activities are considered as evocations of ordered transitions among attractors by some input sequences.

Accordingly, we conceptualize the symbols systems from the viewpoint of dynamical systems under the following correspondences:

- symbols to attractors,
- symbol manipulations to transitions among attractors,
- manipulation rules to order of the transition.

A chaotic neural network is a candidate to implement the symbolic behaviour as a dynamical system. In chaotic neural network, memories are implemented as attractors of the system (Adachi and Aihara, 1997). In some chaotic dynamical systems, transitions among "attractor ruins[2]" have been found (Kaneko and Tsuda, 2003), that is called "chaotic itinerancy". Thus, we may be able to construct a system with plastically learnable symbolic activities by a chaotic dynamical system by introducing the correspondences mentioned above.

3 Model

3.1 coupled NZ map

In this paper, we use a coupled system of chaotic maps, called NZ map (Nozawa, 1992), for a model representing the features of symbols. A single NZ map is a modified version of Hopfield type neuron model. The modification is to add a self-feedback connection and to descritize the time variable using Euler method. Chaotic behaviour is observed in some parameter region, as shown later. The coupled NZ maps system with full connection is a model of module structures of a neural network in the brain. It is known that the coupled system chaotically itinerates among attractor ruins (Nozawa, 1992).

The NZ map, a single element of the proposed system, is given by the following equations:

$$p_i(t+1) = F_{q_i(t)}\{p_i(t)\} \ , \tag{1}$$

$$q_i(t) = -\frac{1}{T_{ii}}\left\{\sum_{j\neq i}^{N} T_{ij}p_j(t) + I_i\right\} \ , \tag{2}$$

$$\begin{aligned} F_q(p) = \; & rp + (1-r) \times \\ & \left[1 - \frac{1}{2}\left\{1 + \tanh\left(\frac{p-q}{2\beta}\right)\right\}\right] \ , \end{aligned} \tag{3}$$

where

- $p_i(t)$: the internal buffer of the ith element at time t, which develops according to the map, $F_{q_i(t)}$, defined by Eqn.(3)
- $q_i(t)$: the activity of the ith element at time t, which includes the influence from the other elements,
- N: the number of elements,
- T_{ij}: the connection coefficient between, the ith and the jth elements,
- T_{ii}: the coefficient of self-feedback connection of the ith element,
- I_i: the threshold of the ith element,
- r, β: the parameters for the shape of the map $F_q(p)$.

[2]An attractor ruin is a region in state space of a dynamical system, in which orbit stays for a while like an attractor and escapes from there.

Since the variable $q_i(t)$ behaves as a parameter of the map $F_{q_i(t)}$, the NZ map varies the shape of the function with time according to $q_i(t)$. For example, when $r = 0.7$ and $\beta = 0.006$, there is a stable fixed point, in the unit interval, near $p = 0$ for $q = 0$ and near $p = 1$ for $q = 1$. For $0 < q < 1$, the map has three branches. Since the slope of the middle branch is less than -1 for positive β, expansion occurs when an orbit comes the middle branch and the dynamics often becomes chaotic. Examples of the map and the dynamics for different q values are shown in Fig. 1. In Fig. 1a) and c), the dynamics are chaotic, while it is period two when q is 0.5 (Fig. 1b)).

This dynamic behaviour of the map induces the coupled system to show transitions among attractors. As Eqn.(2) says, the parameter $q_i(t)$ varies through interactions with the values of other elements $p_j(t)$. Even if the dynamics of some elements fall into fixed points, the parameter q of such elements are changed by the other ones, and then the dynamics escape from the fixed points.

3.2 Embedding Attractors

As mentioned above, we consider that symbols relevant to attractors of a dynamical system. Since the model proposed here is basically associative memory model, we can embed several memory states as attractors of the dynamical system (Hopfield, 1984). To do this, the connection weights between elements should be appropriately set or learned correspondingly to patterns to be embedded.

We use the following equation to determine the connection T_{ij} between the ith and jth elements,

$$T_{ij} = \sum_s (2V_i^s - 1)(2V_j^s - 1) \ , \tag{4}$$

where $\boldsymbol{V}^s = (V_1^s, \cdots, V_N^s)$ is a N dimensional vector with elements 1 ("ON") or 0 ("OFF"). This vector represents a memorized pattern.

3.3 Recalled Pattern

We define, according to (Nozawa, 1992), a recalled pattern $\boldsymbol{\phi}(t) = \{\phi_1(t), \cdots, \phi_N(t)\}$ by observing the

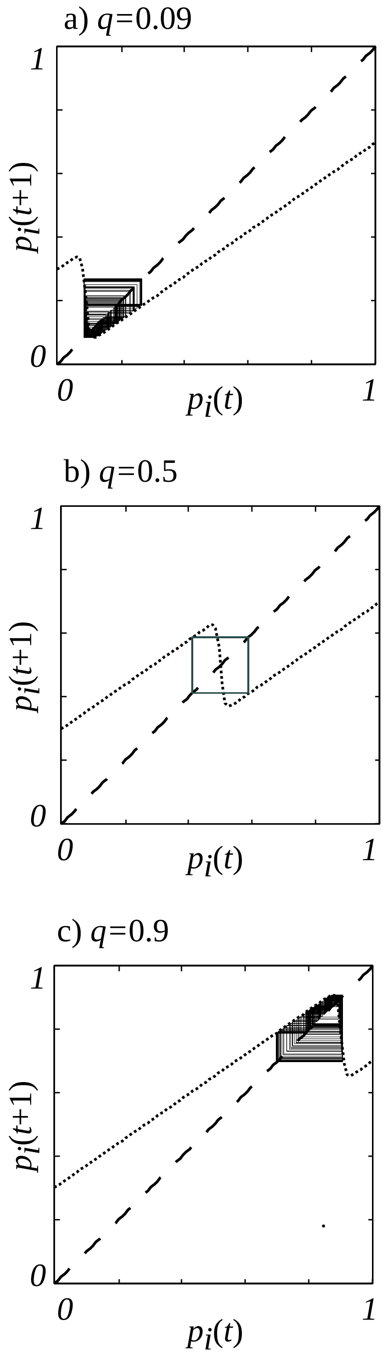


Figure 1: The change of the shape of NZ map $F_q(p)$ and their dynamics for different values of a parameter a) $q = 0.09$, b) $q = 0.5$ and c) $q = 0.9$, when $r = 0.7$ and $\beta = 0.006$. The horizontal and vertical axes are the value of a variable p at time t and $t + 1$, respectively. The dotted lines are the shape of the map. The solid lines shows the dynamics of the map (cobweb plot). The broken lines are diagonals.

values of $q_i(t)$ of all elements as

$$\phi_i(t) = \begin{cases} 1 & (q_i(t) \geq \bar{q}(t)) \ , \\ 0 & (q_i(t) < \bar{q}(t)) \ , \end{cases} \tag{5}$$

where

$$\bar{q}(t) = \lim_{n\to\infty} \frac{1}{tN} \sum_{t'=0}^{t-1} \sum_{i=1}^{N} q_i(t) \tag{6}$$

is the criterion to separate whether each element is "ON" ($\phi_i(t) = 1$) or "OFF" ($\phi_i(t) = 0$). This criterion is the spatiotemporal average of $q_i(t)$. When a recalled pattern is coincide with one of the embedded patterns, that is,

$$\boldsymbol{\phi}(t) = \boldsymbol{V}^s \tag{7}$$

for some s, the embedded pattern, or memory, is retrieved.

4 Simulation Results

We embed three orthogonal patterns, shown in Fig. 2 and named $C, F, 4$, respectively, in the system of 16 elements. As a symmetrical nature of the system, the reversed patterns of the embedded ones are also attractors. Such reversed patterns are labelled as $\bar{C}, \bar{F}$ and $\bar{4}$, respectively. All the other patterns than the embedded and their reversed patterns are treated in a lump and labelled by O. The parameters are $I_i/T_{ii} = 0.09$, $r = 0.7$ and $\beta = 0.006$ throughout this paper.

4.1 Recall and Transition of Embedded Patterns

In a certain region of the strength of self-feedback connection T_{ii}, the system starting from generic initial state recalls one of the embedded or their reversed patterns after some transient behaviour (Fig. 3). The two graphs in Fig. 3 have the same parameters but differ in the initial states. They converge to the different patterns. The fact that different initial conditions end up with the different converged attractor means that this system has multi-attractors.

When we rise the strength of the self-feedback connection from the convergence parameter region, the system itinerates among embedded and non-memorized patterns as shown in Fig. 4. Namely, autonomous changes of recalling patterns are realized in this system.

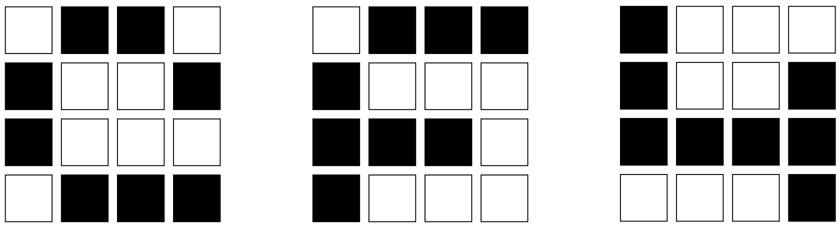


Figure 2: The schematic view of embedded patterns. The black and white boxes mean 1 (“ON”) and 0 (“OFF”), respectively. Each patterns are named as C, F and 4, respectively.

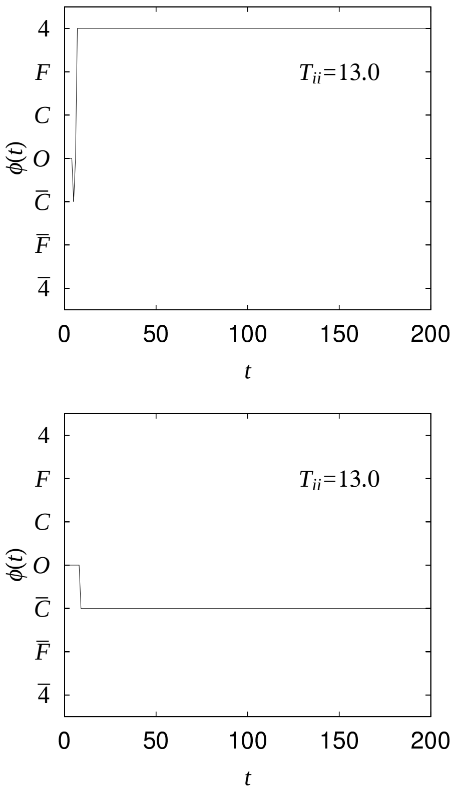


Figure 3: The time series of recalled patterns. The x and y axes are time t and the recalled patterns ϕ, respectively. The labels $C, F, 4$ are the embedded patterns and $\bar{C}, \bar{F}, \bar{4}$ are their reversed patterns, respectively. The label O means that the system is not in any embedded pattern. Two graphs starts from the different initial conditions. The self-feedback connection is $T_{ii} = 13.0$ in the both graphs.

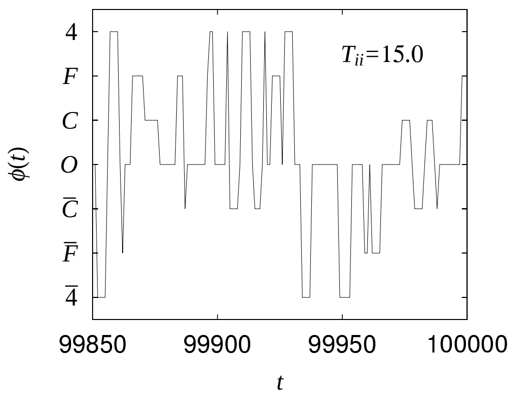


Figure 4: The time series of recalled patterns ϕ without input. The self-feedback connection is $T_{ii} = 15.0$. This graph shows an itinerant motion among the patterns.

4.2 Response to Input

We examine how the system acts to external input. There is a variety of the way to give input for the system. As simple cases, we consider constant and periodic inputs and observe the response of the system to the inputs. The external input $S(t)$ is given as

$$q_i(t) = -\frac{1}{T_{ii}} \left\{ \sum_{j \neq i}^{N} T_{ij} p_j(t) + S(t) + I_i \right\} \quad (8)$$

in Eqn.(2)

In order to observe the dynamics of the system more precisely than the sequence of recalled patterns, we introduce a distance measure of orbit $q_i(t)$ from the embedded patterns $\boldsymbol{V}^{s'}$ [3]. The measure is defined

[3] The index s' is for the patterns both embedded and their reversed.

as

$$Dist^{s'}(t) = \sqrt{\sum_{i=1}^{N}(V_i^{s'} - q_i(t))^2} \ . \qquad (9)$$

4.2.1 Constant Input

We give a constant input sequence $S(t) = 0.3$ at $t = 10000 \sim 20000$ for the system with the self feedback connection $T_{ii} = 15.0$, that is, showing the itinerant motion as shown in Fig 4. By this constant input, the orbit is sometimes fixed at a pattern and sometimes fluctuates among the patterns. Figure 5 shows the time series of $q_i(t)$ around the period of inputting when the system fall onto a fixed pattern responding to an constant input. Only two of 16 elements are depicted. The elements fluctuate largely, actually itinerate among attractors, before the input, and stabilized by the input.

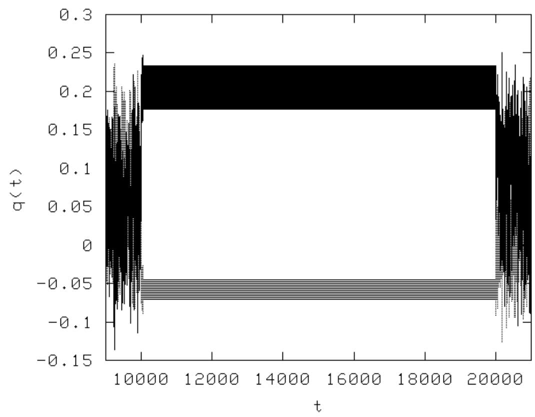


Figure 5: The time series of $q_i(t)$ around the period of inputting ($t = 10000 \sim 20000$) when a fixed input $S(t) = 0.3$ is given. The x axis is time. The self-feedback connection is $T_{ii} = 15.0$. The time series of 2 elements among 16 are drawn.

Figure 6 is a magnification of the inputting period. As this graph shows, the elements compose two clusters according to the range of the values. One cluster moves between 0.2 and 0.3 and the other between -0.05 and 0. While the elements are not fixed as their dynamics of $q_i(t)$, the recalled pattern does not change. The time series of the distance measure, depicted in Fig. 7, tells that the system stays at a state where the nearest pattern is $\bar{F}$. Namely, the elements in the upper cluster in Fig. 6 are "ON", the others are "OFF". The pattern of "ON" and "OFF" coincides with that of $\bar{F}$.

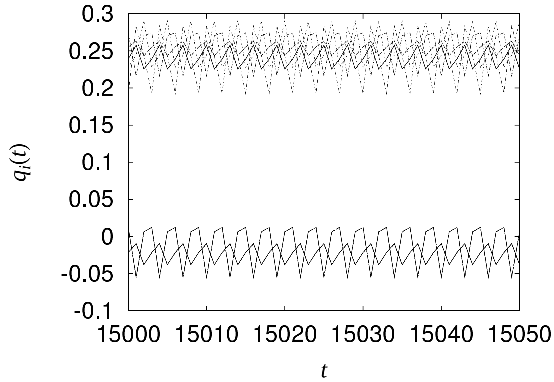


Figure 6: The time series of $q_i(t)$ when a fixed input $S(t) = 0.3$ is given. The x axis is time. The self-feedback connection is $T_{ii} = 15.0$. The time series of all 16 elements are superimposed. They form two clusters.

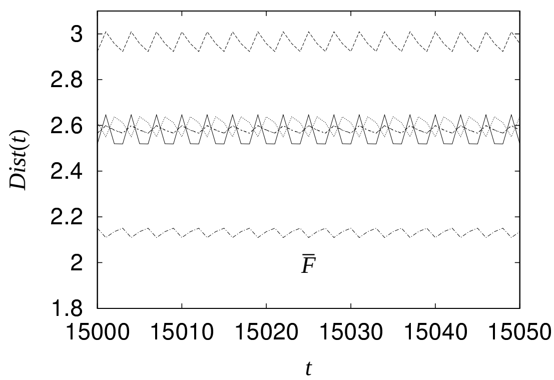


Figure 7: The time series of the distance measure from the embedded patterns, $Dist^{s'}(t)$, when a fixed input $S(t) = 0.3$ is given. The x axis is time. The self-feedback connection is $T_{ii} = 15.0$. The six orbits of the distance from all patterns are superimposed. The orbits composing three clusters periodically change, but the nearest pattern is fixed at $\bar{F}$.

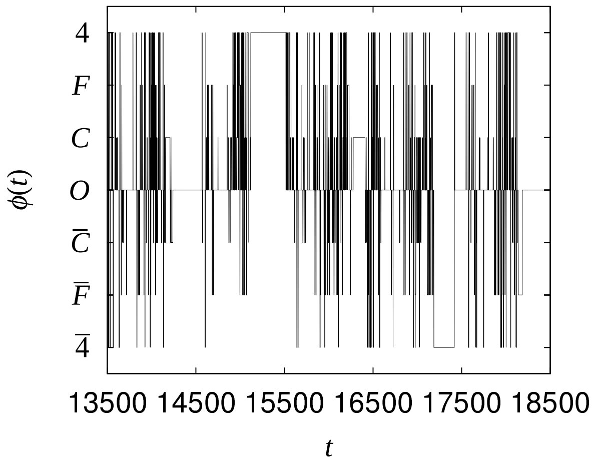


Figure 8: The time series of recalled patterns with a sinusoidal input sequence. The self-feedback connection is $T_{ii} = 15.0$. This graph shows an itinerant motion among the patterns.

Note that while the ON-OFF pattern of the system is the same as the embedded pattern, the values of $q_i(t)$ do not accord with the vectors $\boldsymbol{V}^{s'}$ of the patterns. The attractors of the system with constant input are not fixed point but has dynamics. These facts can be seen clearly in Fig. 7.

As we mentioned, the system with a constant input sequence sometimes converges to a fixed pattern with variety and sometimes itinerates among patterns, in which the itinerant motion is not the same as one without input sequence. This behavioral diversity depends on the timing of the input, since the system is in the itinerant states as shown in Fig. 4. This itinerant behaviour is considered as an internal dynamics of the agent. The agent differs its response to stimuli according to its internal dynamics, even though the same stimulus is given.

4.2.2 Sinusoidal Input

When we input a sinusoidal sequence,

$$S(t) = A\sin(2\pi\omega t) \tag{10}$$

at $t = 10000 \sim 20000$ ($A = 0.7, \omega = 0.001$) to the same system as the previous experiment, the self-feedback connection is $T_{ii} = 15.0$. A transition among pattern is observed as shown in Fig. 8.

For a close observation, we draw the dynamics of the distance measure in Fig. 9. This graph tells us that the change of the nearest patterns occurs with roughly the same intervals. The interval approximately matches with the cycle of the sinusoidal input. The sequence of recalled patterns is not periodic. Further, we have not found the clear statistical order in the transition among recalled patterns.

Thus we can say that the change of the recalled pattern is certainly induced by the periodicity of the sine wave. But when the amplitude of the sine wave comes to around zero, the system is reset to a transient state and is lead into an attractor with growth of the amplitude of the sine wave.

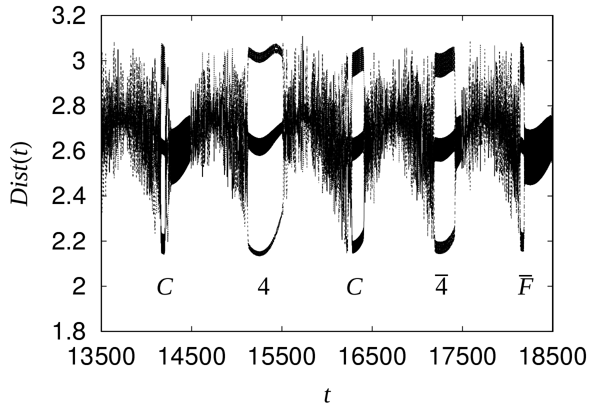


Figure 9: The time series of the distance measure from embedded patters, $Dist^{s'}(t)$, when a sinusoidal input sequence is given. The six orbits of the distance from all patterns are superimposed. The self-feedback connection is $T_{ii} = 15.0$. A transition among the nearest patterns with nearly the same intervals is observed.

5 Discussion

We have summarized the correspondence of symbolic activities to dynamical systems such that symbols correspond to attractors, symbol manipulations to transitions among attractors, and manipulation rules to order of the transition. Let us examine how the simulation results of the coupled NZ maps concordant with these correspondences as the model of a symbolic cognitive agent.

At first, some patterns are embedded to the system and they are retrieved as attractors. The embedded patterns are recalled when a constant input is given to the system. Namely, the patterns and input sequences are associated like memories or concepts and some patterns of perceptions. This signification or representation is an important function of symbols. Further, embedding several patterns or having multi-attractors indicates that the system has some capacity to learn symbols.

Concerning the second point, a dynamical system model of symbolic cognitive agent is required to show, at least, transition among attractors. We can realize such behaviour in some regions of the strength of self-feedback connection T_{ii} as shown in Fig. 4. The transition is evoked by the input sequence of an ordered change. Namely, our system also can have a capacity of the symbol manipulations.

However, the transitions are not orderly. We have not found the clear rule of the transitions, that is, no syntax. We suppose that the reason why no order is shown is partly that the embedded patterns are prepared by hand and are orthogonal. This means that the system does not develop in a particular environment. Namely, there is no structural coupling between the internal structure of the agent and an environment. Learning in a peculiar environment forms the structure of the dynamical system. Thus, to investigate the behaviour of the system leaned in a structural environment is important to examine the systems legitimacy as a model of symbolic cognitive agent.

Let us further discuss the process of symbol formation or development of symbols based on the correspondences between symbolic activities and behaviour of dynamical systems. Harnad (1990) summarizes the developmental process of representation as a progress from iconic representation to categorical and to symbolic. The iconic representation is retrieving an attractor by an input. This is shown by our system.

The categorical representation can be translated in two ways. One is retrieving an attractor for different inputs, the other is grouping of attractors according to some feature of the attractors, such as dimension and nonlinearity. Our system shows the former behaviour. The dynamics of the elements differ for different inputs but they are categorized into one pattern that is nearest to the dynamics. Namely, precise internal states and their dynamics of the agent do not the same but the representing symbol is the same for a class of inputs. To observe the latter interpretation, we need further investigation about the characteristics of attractors of the model.

The symbolic representation is an orderly transition among attractors induced by an input sequences. As we examined, this is not realized in the present system. If we develop, however, the system in a structured environment, the system will learn some symbols as attractors and may show ordered transition among the attractors. Thus, the progressing path of symbolic representation may be able to treated.

6 Conclusion

We have proposed a dynamical system model of cognitive agent that can exhibit a part of symbolic behaviour using a coupled chaotic maps called NZ map. We have shown that symbols as attractors of the dynamical system can be embedded, that the system can have internal dynamics, and that it shows symbols manipulation behaviour as transitions among the embedded attractors according to sequences of external input signals. However, it does not show ordered transitions among symbols, that is, no syntactically structured behaviour.

In spite of this drawback at the status quo, we conclude that the coupled NZ map system can be developed as a model of symbolic individuals, since we may overcome such insufficiency by further study especially on the system learned and developed in a particular structured environment.

Acknowledgements

This research is supported by JAIST research grant (grant for in-house research projects).

References

Masaharu Adachi and Kazuyuki Aihara. Associative dynamics in a chaotic neural network. *Neural Networks*, 10:83–98, 1997.

Angelo Cangelosi and Domenico Parisi, editors. *Simulating the Evolution of Language*. Springer, 2002.

Terrence Deacon. *The Symbolic Species: The Co-Evolution of Language and the Brain*. W. W. Norton, 1997.

Stevan Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

Takashi Hashimoto. The constructive approach to the dynamical view of language. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pages 307–324. Springer, 2002a.

Takashi Hashimoto. Language as dynamics, – a computational study of ontogenetic and glossogenetic loop. In James R. Hurford and Tecumseh Fitch, editors, *Fourth International Conference on the Evolution of Language – Proceedings*, page 51, 2002b.

John J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 81: 3088–3092, 1984.

Kunihiko Kaneko and Ichiro Tsuda. Chaotic itinerancy. *Chaos*, 13(3):926–936, 2003.

Simon Kirby. Learning, bottlenecks and evolution of recursive syntax. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition*, pages 173–203. Cambridge University Press, 2002.

Simon Kirby and James R. Hurford. The emergence of linguistic structure: an overview of the iterated learning model. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pages 121–147. Springer, 2002.

Hiroshi Nozawa. A neural network model as a globally coupled map and applicatoins based on chaos. *Chaos*, 2(3):377–386, 1992.

Ichiro Tsuda. Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences*, 24(5):793–847, 2001.

Tim van Gelder. The dynamical hypothesis in cognitive science. *Brain and Behavioural Sciences*, 10: 615–665, 1998.

Could Navigation Be the Key to Language?

Dimitar Kazakov[⋆]
[⋆]Department of Computer Science,
University of York,
Heslington,
York YO10 5DD, UK,
kazakov@cs.york.ac.uk

Mark Bartlett[†]
[†]Department of Computer Science,
University of York,
Heslington,
York YO10 5DD, UK,
bartlett@cs.york.ac.uk

Abstract

This article analyses navigation and language parsing as two instances of the same abstract computation, and suggests that the tool needed may have evolved to serve the former task, and was then reused for the latter. Supporting evidence for the idea, based on the authors' concept of 'songline' navigation, is discussed in the context of current linguistic, psychological and neuroscience research. The discussion is concluded with an outline of a number of experiments that could shed further light on the subject.

1 Introduction

It is usually assumed that language as used by humans is inherently different from any other form of communication in other species, including primates. Facing this gap, and the intuitive expectation that several mechanisms would have to be already in place before language could play its role and provide any selective advantage, one can question what role (if any) evolution played in the emergence of language. Here we approach this issue by observing that a crucial feature of language, the ability to perform syntactic analysis and generate sentences from a set of grammar rules, can assist navigation, and suggesting that this ability may first have evolved for that purpose, and could be grounded in a general purpose neural circuit performing a certain class of abstract computations applicable across domains.

The article starts with a summary of current theories about the evolution and nature of the *language faculty*, the mechanism that allows us to acquire and use language. The following section discusses the use of computer simulations to model language evolution, and describes the Songlines model of navigation used in our simulations, and its implications on the potential link between the ways humans have evolved to handle navigation and language. This parallel, which we make from an abstract, computational perspective, is then compared with linguistic, psychological and neurological evidence supporting our suggestion. Further, the article describes the design for an experiment aimed at using navigation to detect the subject's ability to perform computation, analogous to context-free parsing. The last section summarises the main ideas and proposes a framework for comparative neuroimaging experiments that could help verify some of the claims made.

2 Evolution and Nature of Language Faculty

While the evolution of languages is an established idea nowadays, emphasised by the practice of grouping languages into genealogical trees of common descent (Pagel, 2000), the quest for the nature of the selective pressure that produced language has not ended in a consensus yet. Language has variously been suggested to emerge in order to provide information about the spacial aspects of the environment (O'Keefe and Nadel, 1978), maintain the social fabric of increasingly larger groups of hominids, e.g., to replace grooming and spread gossip (Dunbar, 1996; Power, 2000), etc. Byrne shows examples of primates using communication as a deception tool (Byrne, 1995), and it can easily be seen how full-fledged language would benefit this ability.

Another related question is the exact nature of the human language faculty and the extent to which it is innate. Chomsky claims that we are born with a Language Acquisition Device (LAD) (Chomsky, 1964), a complex blueprint, which sets its parameters when exposed to language. Marcus et al have reported that seven-month-old infants can learn to discriminate between the sentences of two different grammars (Marcus et al., 1999), reinforcing the belief that this ability is innate rather than acquired. While LAD has many supporters, the claim that it is a prerequisite to using language means that the coexistence of LAD and evolution as leading scientific paradigms is a somewhat uneasy one, as the notion that nature would have to put a potentially very complex tool in place before receiving a payoff contradicts the common wisdom that evolution usually advances in small steps, delivering immediate benefit. A 'macro-mutation' that would have produced the LAD all at once is extremely unlikely, and so would be a hypothesis that the components of LAD have been produced as a series of mutations, each amounting to 'genetic drift', that is, to a change that does not affect one's fitness.

Marcus et al's experiments were based on the familiarisation of subjects with sequences of syllables from an artificial grammar (e.g., both "ga na ga" and "li gi li" are instances of the general pattern ABA). During the test phase, novel spoken sequences, some of which violated the grammar, were played. A strong shift of attention towards the loudspeakers was judged as an acknowledgement of a perceived grammar violation. Importantly, the test sentences consisted entirely of new syllables. The study claims the infants were able to learn to recognise the general ABA pattern as different from ABB. The infants could also discriminate between the patterns AAB and ABB after being familiarised to either.

Marcus et al.'s interpretation of their results is that these are consistent with the infants' being able to "extract abstract algebra-like rules that represent relationships between placeholders (variables)", and that simple statistical learning relying on transitional probabilities cannot account for the experiments' outcome.[1]

Recently, many of the assumptions about the uniqueness of the human faculty of language have started to be questioned and experimentally tested (Hauser et al., 2002). A recent study (Fitch and Hauser, 2004) suggests there is a species (cotton-top tamarin monkeys) that can learn to recognise examples of *spoken* regular languages, but, unlike the human subjects in the studies, failed to learn a context-free language.[2] The results appear to extend the ground humans share with other species, and to point at the ability to handle CFGs as exclusively human.

In the reported work, Fitch and Hauser follow the familiarisation technique used by Marcus et al. (Marcus et al., 1999), but there are differences, for instance in the presence of overlap between syllables used in the training and testing phase. The RG ($(AB)^n$) and CFG (A^nB^n) used by Fitch and Hauser are more complex than those used on infants (ABB and ABA respectively), but still constrain n to 2 or 3 due to the memory limitations of tamarins.

Marcus et al. claim to have honed their methodology to eliminate the chance of having salient features in either grammar that would allow its recognition by statistical means. While following much of their precautions, and putting a considerable effort into eliminating any potentially salient non-grammatical features, Fitch and Hauser's work has been criticised for the use of different speakers for each of the A and B classes of symbols (syllables). If pitch is used as a feature, it is claimed, each of the languages studied would collapse to one example (for each grammar and value of n), thus reducing the experiment to one "about memory span and/or sensitivity to statistical deviations".

3 Songline Navigation

In recent years, there has been much research carried out in attempting to model the evolution of language through computer simulation. This research falls broadly into two classes, simulations in which language emerges in a single generation and simulations concerned with evolving a language over several generations.

In the former class, one of the most prominent researchers is Luc Steels. In his simulations (Steels, 1999), a population manages to arrive at a single, shared lexical language through participating in a series of 'language games'. In a language game, two agents discuss an object visible to both of them. If they can agree on a word (or set of words) to describe that object, then they both increase the weight they associate with that word/meaning pair. After many language games involving different pairs of agents, a shared global lexicon emerges.

Amongst those studying languages which are created over several generations, simulations presented by Kirby (Kirby, 2002) are amongst the most compelling, though Zuidema and Hogeweg (Zuidema and Hogeweg, 2000) and Oliphant and Batali (Oliphant and Batali, 1997) also present interesting results. In Kirby's simulations, a single agent attempts to express (resorting to invention if necessary) a subset of meanings sampled from a set of meanings, expressed in predicate calculus, while another agent attempts

[1] They however rely on the presence of such statistical learning, as observed by Saffran et al. (Saffran et al., 1996) to eliminate an alternative interpretation of their results.

[2] In the article, CFGs are discussed under the more general category of Phrase Structure Grammars.

to learn to speak based on the expressed meanings paired with the linguistic output of the first agent. The agent which listened is then required to speak in the same way as the first agent, while its output is learned by yet another agent. After thousands of cycles of this expression/induction behaviour, a grammar with the minimum number of necessary rules is seen to emerge and persist from generation to generation. Kirby attributes this to the 'linguistic bottleneck' that prevents the observation of all possible meaning/signal pairs by a single agent. Only compositional grammars can successfully pass through this bottleneck, as idiosyncratic phrases present in a grammar may fail to be expressed at some cycle and be lost from the language.

We have also chosen simulations as a means to study the evolution of language, but our approach, first outlined in (Kazakov and Bartlett, 2002), differs from all simulations mentioned above in several important aspects. Primarily, we see language as a tool to achieve some purpose. This means that we can consider issues such as when language will come to be used by a population, whereas other researchers have simply assumed that language is beneficial and sidestepped these issues. So far, our published work also differs by being unconcerned with the evolution of vocabulary: we assume that a shared lexicon has been fixed in the population by some means (for example through language games such as those used by Steels (Steels, 1999)), and concentrate on issues such as the mechanisms by which compositional language may have evolved and the types of environment in which it would be most beneficial (Bartlett and Kazakov, 2004).

Hamilton, and the following neo-Darwinist school of evolution, have developed a formal model of the phenomenon, and demonstrated on numerous examples that sharing among the individuals of a species is compatible with the concept of natural selection in the case of *kin selection* where help is directed to relatives in proportion to the degree of kinship (Hamilton, 1964; Dawkins, 1982). We assume that the mechanism of helping the poorer (hungrier, thirstier) agents has already been established in our simulated society (to simplify the matter, they are all considered equally related), and compare the benefits of sharing information about the location of the resource needed (food, water) between two agents with the case of help in kind, where part of the already collected resource changes hands.

A crucial observation, on which all work is based, is the fact that navigation can benefit from, and be based on a mental representation storing the route be-

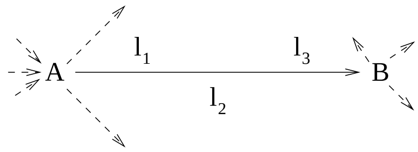


Figure 1: Navigation as a parsing task.

tween two locations A and B as a sequence of landmarks to be passed on the way. In the example in Fig. 1, to go to B, one has to be at (or go to) point A, then pass landmarks l_1, l_2 and l_3, in this order. This can be formally expressed as a rule:

$$\textit{goto}(B) \;::\; \textit{goto}(A) \;\; l_1 \;\; l_2 \;\; l_3 \tag{1}$$

To the experienced eye, this is, of course, a rule of a regular grammar, in which the start and end points of a route play the role of non-terminals (as there may be more than one way to reach and/or leave them), and landmarks are terminals. Therefore, tracing out (or following) a route between two points would amount to generating (resp. parsing) a sentence of a regular grammar (Kazakov and Bartlett, 2004). This representation was inspired by a socio-cultural phenomenon among the Australian Aborigines known as *Songlines*, a form of shared tribal memory, the knowledge of which is mandatory, and often secret, in which each song describes a landmark along a route, and the series of songs constitutes a sung map (Barwick and Marrett, 2003).

All that is needed to exchange a route between two agents using 'songlines' as their internal representation of the environment, is a shared lexicon of landmark names. Steels's experiments show that such a lexicon can easily be evolved from simple first principles (Steels, 1999). We assumed the existence of this lexicon, and sought to identify the types of environment in which sharing songlines outperforms selfish behaviour and sharing in kind. Among the factors studied were the abundance of the two types of resource modelled, and their volatility The results show language is particularly beneficial in the border zone delineated by the trade-off given by the simultaneous increase/decrease of resource availability and volatility. At one side of this line, there is too litle food and the one found disappears too quickly to be worth going back to; on the other side, food is sufficient to permit the survival of selfish agents. The area in the middle is where language appears to make the difference between survival and extinction (Kazakov and Bartlett, subm). In those areas, language outperforms both selfish behaviour and sharing previously accumulated resource.

4 Parallels between Navigation and Parsing

We have demonstrated that memorising and planning routes by an agent that describes a path as a sequence of landmarks (beacons) amounts to storing the rules of a regular language and generating/parsing its sentences. This is important: *if a regular language parser* (i.e., a Finite-State Automaton (FSA)) *could help navigation, it may first have evolved for this purpose*. Then only a relatively small change in the neural connections, possibly even caused by a single mutation, might have been required to make this parser available to the human brain speech circuitry. This compares favourably with the idea of macromutation, as described above. The idea of separately evolved needs for lexicon and grammar are also consistent with evidence that they are separated in the brain (Ullman, 2004).

We can consider now navigation and language parsing as two instances of the same abstract computation (involving strings of symbols) and enquire whether the way we perform these tasks would reflect that. Anyone interested in this question would be likely to look into existing models of the way in which syntax is grounded in the neural substrate. Ullman's recent model (Ullman, 2004) pinpoints several memory circuits in the brain, "a network of specific frontal, basal-ganglia, parietal and cerebellar structures", which support "the learning and execution of motor and cognitive skills, especially those involving sequences". The model separates, both neurophysiologically and conceptually, this so called procedural memory from the declarative memory storing information about facts and events, including the mental lexicon. The suggested common basis for the processing of verbal and non-verbal sequences is supported by other authors. Hoen et al. (Hoen et al., 2003) report that using non-verbal symbols (playing cards) to exersise the ability to reorder sequences in a predefined way ($123 \rightarrow 231$), helps patients improve their ability to understand a type of sentences that need the same transformation to have their constituents rearranged in the default order: "It was the cat^1 that the dog^2 chased^3" $\rightarrow$ "The dog^2 chased^3 the cat^1". Hauser, Chomsky and Fitch (Hauser et al., 2002) also draw a link between navigation and language, suggesting understanding efficient processing of language can help research in other domains, " such as spatial navigation and foraging, where problems of optimal search are relevant".

There are two ways in which the link between motor and verbal sequence processing may hold the key to the origins of syntax. One could conceive two coupled processes, (1) the need for 'songline' navigation providing selective pressure for the evolution of a parser, and, (2) the advantages of sharing 'songlines' promoting language. While it is very tempting to suggest that not only navigation may have selected for the evolution of a parser, but that it may have been the first topic of discussion to which this parser was applied, the second need not be the case. In fact, it is not easy to imagine how such a specialised language, possessing only nouns (or noun phrases) and the single verb '*to go*', would have developed other parts of speech as a function of the navigational task. On the other hand, one can also consider another course of events, that the parser originally evolved to serve navigation as a means of internal representation and planning, and it was at a later stage that this parser (or its replica) became involved in communication. This resembles the idea of virtual mind machines processing whole classes of computationally homologous tasks, proposed by A. Sloman. This idea seems to be mirrored in neurological evidence about the parallel circuits of basal ganglia as performing analogous computations, applied to different sets of information from different domains (Ullman, 2004).

5 From Regular to Context-Free Languages

So far, we have only demonstrated that navigation could have created the need to represent and process regular languages. However, RGs are not sufficient to describe the seemingly simple task of going somewhere and returning back using the same way. In terms of 'songlines', the landmarks passed on the round trip would spell out a palindrome (e.g., $abcba$ or $aabbaa$). It is known that a parser that can recognise a palindrome, can handle any CFL. A CFL parser is usually modelled as a push-down automaton, consisting of a FSA (i.e., a RL parser) and memory (e.g. stack). Regardless of whether this task created the original need for CF parsing or not, it can be used in an experiment involving navigation, rather than responses to speech, which could help avoid the criticism to Fitch and Hauser's work (Fitch and Hauser, 2004), while still assessing the subject's ability to learn syntax.

One could consider two classes of tasks. In the first, the subject would have to learn to navigate be-

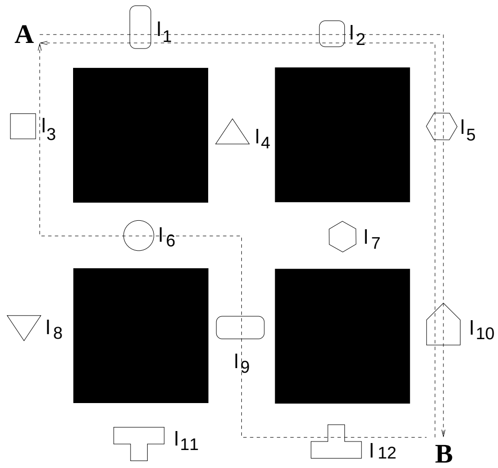


Figure 2: Regular vs Context-Free Songlines.

tween two locations, e.g., A and B in figure 2, collecting a reward each time either location is reached. Assuming landmarks are used to memorise the paths between A and B as suggested above (Kazakov and Bartlett, 2004), successful navigation would amount to learning a regular grammar, e.g.:

$$\textit{Reward} \rightarrow B \quad (2)$$
$$B \rightarrow A\ l_1\ l_2\ l_5\ l_{10} \quad (3)$$
$$\textit{Reward} \rightarrow A \quad (4)$$
$$A \rightarrow B\ l_{12}\ l_9\ l_6\ l_3 \quad (5)$$

The second experiment modifies the above setting by extending the reward given in location A to a more significant one, provided the subject went from A to B and back using the same path. Navigation based on a regular grammar with alternative routes will fail to collect the extended reward most of the time, e.g., there are 18^2 ways of going from A to B and back in figure 2, but only 18 of these will bring the maximum reward. However, a simple context-free grammar will suffice:

$$\textit{Reward}_1 \rightarrow A\ \textit{Reward}_2\ A \quad (6)$$
$$\textit{Reward}_2 \rightarrow X \quad (7)$$
$$X \rightarrow \textit{Landmark}\ X\ \textit{Landmark} \quad (8)$$
$$X \rightarrow B \quad (9)$$
$$\textit{Landmark} \rightarrow l_i,\ 1 \leq i \leq 12 \quad (10)$$

Using another string of salient features, such as turns and distances will not change the need to reverse that string to navigate back (but will assume the ability to transform a left turn to a right one and vice versa). One can use obstacles to guarantee that no complete path used in training is available in the test phase and vice versa.

Classical Reinforcement Learning (RL) (Sutton and Barto, 1998) cannot account for learning this navigational behaviour. With consecutive phases of (random walk) exploration and exploitation, RL will assign in average equal rewards to all A *to* B, resp. B *to* A paths; if the agent alternates between exploration and exploitation, and gradually increases the latter, the earlier an A *to* B, resp. B *to* A path is discovered, the more likely it is to be subsequently reinforced, and given preference in the long term.

The above setting would avoid issues stemming from the much greater importance speech has for humans and put them on a more equal ground with other species. A confirmation of Fitch and Hauser's conclusions that possessing a CF parser is a distinctly human feature would raise the question whether this did not initially evolve to serve non-linguistic purposes, such as navigation, and study the circumstances that made this new feature evolutionary beneficial. For instance, taking the same way back home may reduce the risk of encountering unexpected dangers or help estimate the time needed to return back. On the other hand, a shared need for navigation among species would be consistent with them sharing the ability to learn regular languages.

6 Discussion

The main ideas in this article can be summarised as follows:

1. The ability to hangle regular grammar, a critical step on the road to human language, may originally have evolved to assist navigation.

2. The shared need for navigation should be mirrored in the ability of other species to learn regular languages.

3. Navigation and language parsing are two instances of the same abstract computation, and the way they are grounded may reflect that.

4. The need for context-free grammars, typical for human languages, could have originated in navigation.

The range of indirect evidence for the above statements suggests the idea of using neuroimaging to compare the brain activity between tasks correspongmarks ing to regular and context-free languages for navigation on one hand, and language, on the other. An exciting, but yet unconfirmed possibility is that the patterns of activation for navigation and language would be similar for the same class of languages, but processing a different class of language would result in distinguishable differences even for the same type of task.

References

M. Bartlett and D. Kazakov. The role of environment structure in multi-agent simulations of language evolution. In *Proceedings of the Fourth Symposium on Adaptive Agents and Multi-Agent Systems (AAMAS-4)*, Leeds, 2004. AISB Convention.

L. Barwick and A Marrett. *Currency Companion to Dance and Music in Australia*, chapter Aboriginal Traditions, pages 26–28. Currency Press, 2003.

Richard W. Byrne. *The Thinking Ape: Evolutionary Origins of Intelligence*. Oxford University Press, 1995.

Noam Chomsky. *Current Issues in Linguistic Theory*. Mouton, The Hague, 1964.

R. Dawkins. *The Extended Phenotype*. Oxford University Press, 1982.

Robin Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1996.

M. T. Fitch and M. D. Hauser. Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 2004.

W. D. Hamilton. The genetic evolution of social behaviour I, II. *Journal of Theoretical Biology*, 7: 1–52, 1964.

M. D. Hauser, N. Chomsky, and M. T. Fitch. The faculty of language. *Sci.*, 298, 2002.

M. Hoen, M. Golembiowski, E. Guyot, V. Deprez, D. Caplan, and P. F. Dominey. Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport*, 14, 2003.

D. Kazakov and M. Bartlett. A multi-agent simulation of the evolution of language. In *Proceedings of Information Society Conference IS'2002*, pages 39–41, Ljubljana, Slovenia, Oct 2002. Morgan Kaufmann.

Dimitar Kazakov and Mark Bartlett. Cooperative navigation and the faculty of language. *Applied Artificial Intelligence*, 19, 2004.

Dimitar Kazakov and Mark Bartlett. Benefits of using language to share information in a dynamic environment. Submitted to Connection Science (special issue on Evolution of Language), subm.

S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, 2002.

G. F. Marcus, S. Vijayan, S. Bandi Rao, and P. M. Vishton. Rule learning by seven-month-old infants. *Science*, 283, 1999.

J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Oxford Univ. Press, 1978.

M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language*, 11(1), 1997.

M. Pagel. The history, rate and pattern of world linguistic evolution. In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge, 2000. Cambridge University Press.

C. Power. *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, chapter Secret language use at female initiation: Bounding gossiping communities. Cambridge University Press, Cambridge, 2000.

J. Saffran, R. Aslin, and E. Newport. Statistical learning by 8-month old infants. *Science*, 274, 1996.

Luc Steels. The spontaneous self-organization of an adaptive language. In Koichi Furukawa, Donald Michie, and Stephen Muggleton, editors, *Machine Intelligence 15*, pages 205–224. Oxford University Press, 1999.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

M.T. Ullman. Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 2004.

W. H. Zuidema and P. Hogeweg. Selective advantages of syntactic language - a model study. In *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*, pages 577–582, Hillsdale, USA, 2000. Lawrence Erlbaum Associates.

The Evolution of Meaning-space Structure through Iterated Learning

Simon Kirby*
*Language Evolution & Computation Research Unit
University of Edinburgh
40, George Square, Edinburgh EH8 9LL
simon@ling.ed.ac.uk

Abstract

In order to persist, language must be transmitted from generation to generation through a repeated cycle of use and learning. This process of *iterated learning* has been explored extensively in recent years using computational and mathematical models. These models have shown how compositional syntax provides language with a stability advantage and that iterated learning can induce linguistic adaptation. This paper presents an extension to previous idealised models to allow linguistic agents flexibility and choice in how they construct the semantics of linguistic expressions. This extension allows us to examine the complete dynamics of mixed compositional and holistic languages, look at how semantics can evolve culturally, and how communicative contexts impact on the evolution of meaning structure.

1 Introduction

One of the most striking aspects of human linguistic communication is its extensive use of compositionality to convey meaning. When expressing a complex meaning, we tend to use signals whose structure reflects the structure of the meaning to some degree. This property is the foundation upon which the syntax of language is built. It is natural, therefore, that an evolutionary account of human language should contrast compositional communication with a non-compositional, holistic alternative. Indeed, Wray (1998) has argued that holistic communication (which is still in evidence in particular contexts today) can be seen as a living fossil of an earlier completely non-compositional protolanguage.

A compositional syntax has clear adaptive advantages – with it we are able to successfully communicate novel meanings (in the sense that we may never have witnessed signals for those meanings in the past). Despite this, research over the past decade has suggested that compositional syntax may have emerged not because of its utility to us, but rather because it ensures the successful transmission of language itself (see e.g. Kirby, 2000). It is suggested that the process of linguistic transmission, termed *iterated learning* (Kirby & Hurford, 2002), is itself an adaptive system that operates on a timescale intermediate between individual learning and biological evolution. Computational models of this process (e.g. Kirby, 2000; Batali, 1998) have demonstrated that syntactic systems can emerge out of random holistic ones without biological evolution, at least for particular assumptions about learning, production and so on.

Further evidence for the argument that iterated learning can explain features of syntax has been provided by idealised computational (Brighton & Kirby, 2001) and mathematical (Brighton, 2002) models of iterated learning in general showing that compositional languages have a stability advantage over holistic ones. These models compare two scenarios under a number of different parameters. They analyse completely holistic languages and completely compositional ones. The parameters that are varied relate to, on the one hand, the structure of the meaning space, and on the other, the number of training examples an individual is exposed to (also known as the *bottleneck* on linguistic transmission). The overall conclusion is that with highly structured meaning spaces and few training examples, compositional languages are more stable than holistic ones.

2 Problems

This foundational work on the cultural evolution of meaning-signal mappings through iterated learning, though important in demonstrating that language itself has significant adaptive dynamics, suffers from two significant drawbacks, which we will turn to below.

2.1 Stability analysis

Early models such as Batali (1998) and Kirby (2000) involved populations of individual computational agents. These agents were equipped with: explicit internal representations of their languages (e.g. grammars, connection weights etc.); a set of meanings (provided by some world model) about which they wished to communicate; mechanisms for expressing signals for meanings using their linguistic representations; and algorithms for learning their language by observing meaning-signal pairs (e.g. grammar induction, back-propagation etc.).

Typically, these simulations initialise the population with no language, or a random pairing of meanings and signals and then allow the linguistic system to evolve through repeated encounters between speaking agents and learning agents.

There has been much work in building simulation models within this general iterated learning framework (e.g. Batali, 1998; Kirby, 2000; Tonkes, 2001; Kirby & Hurford, 2002; Brighton, 2002; K. Smith, 2003; Zuidema, 2003). The great advantage of this kind of modelling is that it allows the experimenter to demonstrate possible *routes* by which language can evolve from one qualitative state, such as holistic coding, to another, such as compositionality.[1] The models show how fundamental features of language can emerge in a population over time given reasonable assumptions about how linguistic behaviour may be transmitted.

Models such as these tend to have a large range of parameters, and it is therefore reasonable to want to know the relationship between the emergent property and the parameter space of the model. Once we understand this, we can eventually hope to uncover theoretical principals that may apply to iterated learning *in general* rather than the specific model in question.

As mentioned above, two key parameters in the emergence of compositionality are: meaning-space structure (i.e. the set of things agents communicate about); and learning bottleneck[2] size (i.e. the number of training examples agents are exposed to).

Computational simulations indicate that it is important that there is some kind of learning bottleneck for there to be any interesting linguistic evolution. To put it simply, only when training data is sparse will language evolve to be compositional.

[1] The emergence of compositionality has received a lot of attention. However, it is important to note that other fundamental linguistic universals may well explicable within this general framework. The central message is that wherever there is iterated learning, there is potential for adaptation of the system being transmitted to maximise its own transmissibility.

[2] See Hurford (2002) for discussion of why the term "bottleneck" is appropriate, and for an analysis of different types of bottleneck in language evolution.

This parameter is relatively straightforward to experiment with, but meaning-space structure is far more difficult, and most of the simulations of iterated learning simply chose some kind of system of meaning representation and stuck with it for all simulations.

The work of Brighton & Kirby (2001) and Brighton (2002) was an attempt to get round this problem by exploring a large range of possible meaning-spaces and examining what impact they would have in an iterated learning model.

In those papers – as in this one – a highly idealised notion of "meanings" is employed: meanings are simply feature vectors. A meaning-space is defined by the number of features F it has and the number of different values V over which each feature can vary. So, to communicate about a world where objects were either squares, circles or triangles, and could be coloured green, blue or red, agents would need a meaning-space with at least $F=2$ and $V=3$.

A reasonable strategy for thoroughly exploring the role of meaning-space structure might be to run many iterated learning simulations, each with a different meaning space, and determine the trajectory of the linguistic system in each instance. This proves computationally costly, so Brighton and Kirby instead looked at what would happen to either a completely compositional language or a completely holistic one for each meaning-space.

Firstly using a computational model, and then using a mathematical generalisation of this model, they were able to calculate how stable either language type was for all meaning spaces. Simplifying somewhat, the overall result was that compositional languages have a stability advantage over holistic ones for larger meaning spaces, especially where the number of features are high.

This kind of simplification of the iterated learning process is very useful but leads to the first of our two problems. Whereas a standard iterated learning simulation can demonstrate a trajectory, or route, from holism to compositionality, the Brighton and Kirby idealisation can only tell us about the relative stability of end-points of such a trajectory. In other words, we don't know whether there is a way to get to a stable compositional language from an unstable holistic one because we don't know anything about the languages in-between.

2.2 Fixed, monolithic meaning space

A second problem with much research into iterated learning so far has been its reliance on a pre-existing meaning space provided for and shared by all agents in the simulation.[3] The work described in the previous section makes strong claims about the likelihood of the emergence of compositional syntax given a particular prior space of meanings. But, where does this meaning space come from? It is assumed that biological evolution somehow endows the agents with a representational scheme prior to language, and if those representations are of sufficient complexity, a compositional system of expressing them will follow naturally.

Furthermore most, if not all, models assume that there is a single, monolithic system for representing meanings. Everything the agents in the simulations want to talk about can be expressed in the same format, be that a feature vector of particular dimensionality, a predicate-logic representation, or a point on a real-number line etc. Equally, there is assumed to be one and only one meaning for representing every "object" in the agents' world.[4]

As with the study of the relative stability of "end-points" in language evolution, a monolithic, fixed and shared meaning-space is a sensible idealisation to make. Modellers hold one aspect of the object of study constant – meanings – and allow another aspect – signals – to evolve through iterated learning. Much has been learned through these idealisations, but equally it is important to explore what happens if we relax these assumptions.

3 A simple model

In this paper I will set out a simple extension to the model in Brighton (2002) which allows us to look at what happens when agents have flexible meaning representations for objects. It turns out that this extension also allows us to move beyond a simple stability analysis of end-points of iterated learning and give us, for the first time, a complete view of the dynamics of iterated learning.

3.1 Meanings

Language can be viewed as a system for mapping between two interfaces (see, e.g., Chomsky, 1995). On the one hand, there is an articulatory/perceptual interface, which handles input and output of signals. On the other, there is a conceptual/intentional interface, which relates linguistic representations to the things we actually communicate about. It is primarily the latter of these two that we are concerned with here.

In the model, there is a predefined set of things about which the agents wish to communicate – we will call this the *environment*, *E*. The conceptual/intentional interface *C* consists of a number of *meaning spaces* $M_{\langle F,V \rangle} \in C$ onto which every object $o \in E$ in the environment is mapped. Each of these meaning spaces, in keeping with previous models is defined as a set of feature-vectors, such that each meaning space is defined by the number of features *F* it has (its *dimensionality*), and the number of values *V* each of these features can take (its *granularity*).

Throughout a simulation run, every object in the environment is paired with a particular point in every meaning space. For the simulation runs described here, this is set up completely randomly at the start of the run. Loosely speaking, we can think of this as giving an agent a number of different ways of conceiving an object. Note that each point in each meaning space can be mapped to zero, one or many objects in the environment. So, for example, there may be particular feature-vectors in particular meaning spaces that are *ambiguous* in that they map to more than one object in the environment.

The important point here is that agents are prompted to produce expressions for *objects in the environment* and not meanings themselves. Part of the task of the agent is to choose which of that object's meanings will be used to generate the linguistic expression. It is this that is the novel extension to previous work. Previously, only one meaning-space was available, so expressing an object and expressing a meaning were the same thing. Now that the latter is under the control of the agent the use of meanings can be learned and, ultimately, itself be subject to cultural evolution through iterated learning.

3.2 Learning

In this model I will follow Brighton (2002, 2003) in considering the task of learning a compositional system to be one of memorising signal elements that correspond to particular values on particular features. A single compositional utterance carries information about how to express each feature-value of the meaning expressed by that utterance.

If we consider just a single meaning space, then learning a perfect compositional system proceeds exactly as in Brighton (2002, 2003). The learner is exposed to a series of *R* meaning/signal pairs

[3] This is not true of the extensive work on symbol grounding carried out by, for example, Steels & Vogt, 1997; Steels, 1998; A.D.M. Smith, 2003; Vogt, 2003.

[4] The term "object" is used here by convention to stand-in for any communicatively relevant situation. In other words, an "object" is anything that an agent may wish to convey to another agent through language.

$(p_1, p_2, \ldots, p_R)$ each of which represents a point in the space $F \times V$. After this exposure, the learner is able to express at least as many meanings as are uniquely expressed in the training data. Note that this is likely to be less than *R* since meanings may be repeated.

Is this the best expressivity that the learner can expect to achieve after learning? Not if the learner is exposed to a compositional language. The learner may be able to express novel combinations of feature-values as long as each feature-value occurs somewhere in the training data.

Brighton (2003) gives the following simple approach to modelling the transmission of a compositional language. The first step is to construct a lookup table recording how each feature-value is to be expressed. This table, O, is an $F \times V$ matrix of signal elements. In fact, in this model the actual nature of those signal elements is irrelevant. This is based on the assumption that the learner can correctly generalise a compositional language from the minimum exposure. Brighton terms this the *assumption of optimal generalization*. (This idealises away from the task of decomposing the input signal into parts and identifying which parts of the signal correspond to which parts of the meaning. We should be aware that, in a more realistic scenario, more data is likely to be required and furthermore, segmentation errors are likely to occur.)

The benefit of this assumption is that we can simply treat each entry in the *O* matrix as a truth value:

$$O_{i,j} = \begin{cases} \textbf{true} & \text{if the jth value of the ith feature is observed} \\ \textbf{false} & \text{otherwise} \end{cases}$$

When the entry $O_{i,j}$ is true, this means that the sub-signal for the j^{th} value of the i^{th} feature has occurred at some point in the training data.

On receiving some meaning/signal pair $p = \langle m, s \rangle$ the matrix is updated so that each of the feature-values contained in *m* are logged in the *O* matrix. If $m = (v_1, v_2, \ldots v_F)$, then:

$$O_{i,v_i} = \textbf{true} \quad \text{for } \mathrm{i} = 1 \text{ to F}$$

So far, this is simply a restatement of Brighton's (2003) formalism. The novel feature here is just that there are multiple meaning-spaces, and therefore multiple *O* matrices to keep track of. To simplify matters for this paper, we will maintain the assumption that learners are given meaning-signal pairs. That is, learners are able to infer which point in which meaning-space a speaker is expressing. It is a topic of crucial and ongoing research, particularly by those researchers looking at symbol-grounding, to develop strategies to relax this assumption (e.g., Steels & Vogt, 1997; A.D.M. Smith, 2003).

So far, contra Brighton (2002, 2003), we have not looked at *holistic* languages. Holistic languages are those where meanings are unanalysed and each given distinct, idiosyncratic signals. Learners cannot, therefore, generalise beyond the data that they are given. However, we can simply equate a holistic language with a compositional language for a meaning-space with only one feature. The machinery described so far, is therefore sufficient to explore the difference between compositional and holistic language learning – we simply need to provide agents with the relevant meaning-spaces.

3.3 Language production

We have specified an **environment** containing **objects** each of which are labelled with **feature-vectors** drawn from each of a set of **meaning-spaces**. We have set out a model of learning whereby sets of **meaning-signal pairs** given to a learning agent are transformed into **O matrices**, one for each meaning-space.

In order to complete a model of iterated learning, it is necessary to provide agents not just with a way of learning, but also a way of producing behaviour for future generations of agents to learn from.

Clearly, a particular meaning $m = (v_1, v_2, \ldots v_F)$ can be expressed by an agent if, and only if, that agent has a way of expressing each feature-value using the language it has learned so far. In other words, iff $O_{1,v_1} \wedge O_{2,v_2} \wedge \ldots \wedge O_{F,v_F}$.

It is important to note, however, that the agents in this model are not prompted to express a *meaning*. Rather, they attempt to produce expressions for *objects* in the environment. This means that an agent may have a choice of potential meaning spaces to employ when signalling about any one object. An object is expressible, therefore, if *any* of the meanings associated with that object are expressible. If more than one meaning is expressible by an agent, a choice must be made. For the first simulations described below, that choice is simply made at random.

The goal of language production in this model is to produce a meaning-signal pair. However, learning as described in the previous section actually makes no use of signals because of the assumption of optimal generalisation. This means we can ignore the signal part of the signal-meaning pair. When a learning agent observes the behaviour of a speaker, the simulation need only note the set of meanings used.

3.3 Simulation run

A simulation run consists of the following steps:

1. **Initialise environment.** Associate each object in the environment with a single random meaning in every meaning space.
2. **Initialise population.** In this simple model, the population consists of a single speaker, and a single learner. At the start of the simulation, the *O* matrices of the adult speaker are initialised with patterns of "true" and "false". The particular way in which they are filled depends on the experiment being run, and represents the initial language of the simulation. The learner's *O* matrices are filled uniformly with "false" because learners are born knowing no language.
3. **Production.** An object is picked randomly from the environment. A list of candidate meanings – one from each meaning space – is compiled for the object. The O matrices of the speaker are used to determine which, if any, of these candidates the speaker can express. One of these is picked at random.
4. **Learning.** If the speaker has been able to find an expressible meaning, the learner takes that meaning and updates its own *O* matrix for that meaning space.
5. **Repeat.** Steps 3 and 4 are repeated *R* times (this defines the size of the learning bottleneck).
6. **Population update.** The adult speaker is deleted, the learner becomes the new speaker, and a new learner is created (with *O* matrices filled with "false" entries).
7. **Repeat.** Steps 3 to 6 are repeated indefinitely.

The relevant simulation parameters are: size of bottleneck, *R*; number of objects in the environment, *N*; the make-up of the conceptual/intentional system, *C* (i.e. the particular $\langle F,V \rangle$ values for each $M_{\langle F,V \rangle}$); and the initial language (i.e. the *O* matrices for each meaning space in *C*).

4 Results

This simulation model can be used to explore the dynamics of iterated learning given multiple meaning-spaces. Because, as mentioned earlier, holistic languages are identical to compositional languages for 1-dimensional meaning-spaces, it can also be used to examine how compositional communication can arise out of a prior holistic protolanguage.

4.1 Meaning space stability

As many previous models have shown, compositional languages are more stable than holistic ones through iterated learning with a bottleneck. We can track *expressivity* of the agents' languages in a simulation over generations given an initial completely expressive language that is compositional, and compare that with a simulation initialised with a completely expressive language that is holistic.

iteration	0	1	2	3	4	5	6	7	8
holistic	1	.45	.22	.13	.08	.02	.02	.02	0
comp.	1	1	1	1	1	1	1	1	1

This table shows expressivity (as a proportion of all the objects in the environment) over time for a simulation with $N = 100$, $R = 50$, $C = \{M_{\langle 8,2 \rangle}\}$ and a simulation with $N = 100$, $R = 50$, $C = \{M_{\langle 1,256 \rangle}\}$.

Unsurprisingly, the holistic language cannot survive in the presence of a bottleneck. The size of the bottleneck affects the rate of decay of expressivity in the holistic language:

iteration	0	50	100	150	200	250	300
R=100	1	0	0	0	0	0	0
R=200	1	.15	.1	.06	.06	.04	.02
R=300	1	.3	.21	.16	.16	.16	.12
R=400	1	.61	.43	.38	.34	.32	.31

As in previous models, this demonstrates once again the crucial advantage a language gains from a compositional syntax.

4.2 Complete holistic/compositional dynamics

Recall that one of the motives for this extension to previous work to move beyond simple stability analysis to see the complete dynamics of the move from holism to compositionality. To do this, we can simply run simulations with two meaning spaces instead of one, such as: $C = \{M_{\langle 8,2 \rangle}, M_{\langle 1,256 \rangle}\}$.

A particular point in the space of possible languages can be described in terms of the proportion of objects that can be expressed using the compositional language, $M_{\langle 8,2 \rangle}$ and the proportion of objects that can be expressed using the holistic language, $M_{\langle 1,256 \rangle}$.

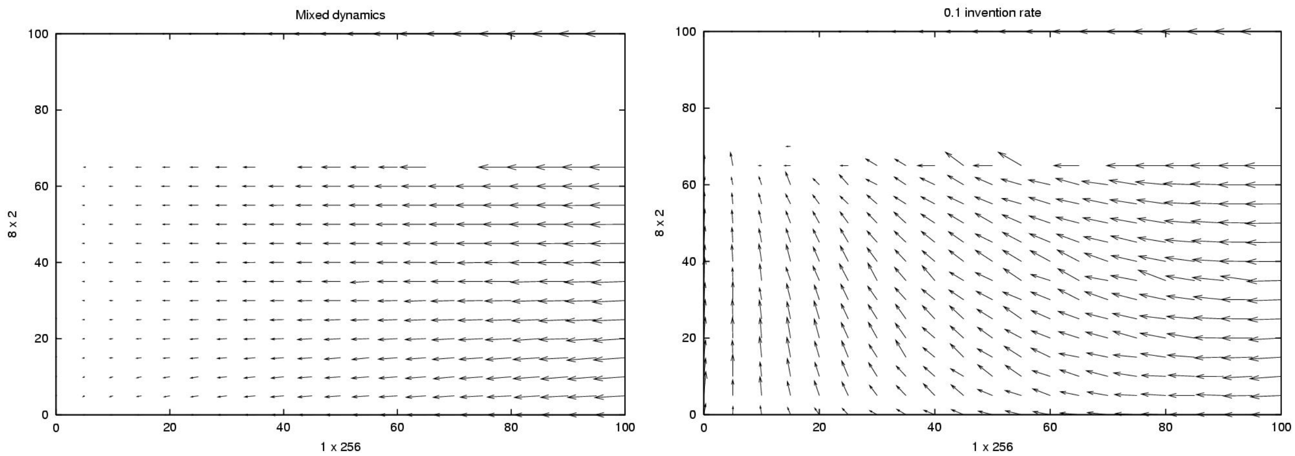


Figure 1a,b: Complete dynamics for languages that are partially holistic and partially compositional, without invention and with invention. Each point represents a language with a particular combination of holistic and compositional signals. Each arrows show the direction and magnitude of movement in this space after a single instance of learning, and represents the average of 100 simulation runs. (The gaps in the graph result from points in this space that cannot be constructed for an environment of 100 objects.)

The complete dynamics for all points in holistic/compositional space is visible in figure 1a. The arrows show the magnitude and direction of change after one iteration of the model for that particular combination of holistic versus compositional expressivity. There is a single attractor at (0,0). In other words, the inevitable end state is one where no objects are expressible either holistically or compositionally.

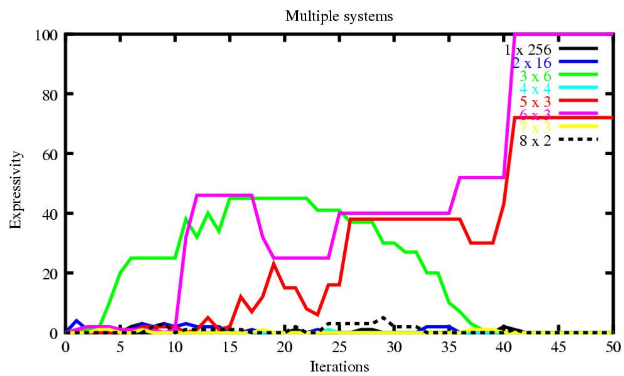
The reason for this is obvious: once a word is lost from the language, there is no way of getting it back. In fact, the agents rely on the expressivity of the language that is injected at the start of the simulation. To get round this, most iterated learning models allow agents to "invent" new expressions. To model this, a new parameter is added – the invention rate *I*. This gives the probability that, on failure to find any way of expressing an object, an agent will pick a meaning space at random and invent an expression for the relevant meaning in that space.

Figure 1b shows how an invention rate of $I = 0.1$ affects the dynamics of iterated learning. Now, the single attractor is the completely compositional language. This demonstrates that there is a clear route from all parts of the language space towards a completely compositional language, through intermediate mixed languages.

As has been shown before, the size of bottleneck is a crucial determinant of whether compositionality will replace holism. If the size of the bottleneck is increased, holistic utterances no longer have a disadvantage and the movement to the left-hand side of these plots is removed. It is the fact that language must pass through a learning bottleneck as it is transmitted from generation to generation that causes it to adapt and causes idiosyncratic non-compositional expressions to die out.

4.3 The evolution of meaning spaces

The second motivation for the current model was to see how iterated learning might result in adaptation of the meanings of expressions as well as the form of the expressions themselves. Previous models used a monolithic, fixed meaning space, but the current model allows for any number of meaning spaces to exist concurrently. An agent's learning experience (and hence, ultimately, its cultural inheritance) decide the structure of the meaning used to express an object in the environment.



The graph above shows an example simulation run with the following initial parameters:

$I = 0.1,\ N = 100,\ R = 50,$
$C = \{M_{\langle 1,256\rangle}, M_{\langle 2,16\rangle}, M_{\langle 3,6\rangle}, M_{\langle 4,4\rangle}, M_{\langle 5,3\rangle}, M_{\langle 6,3\rangle}, M_{\langle 7,3\rangle}, M_{\langle 8,2\rangle}\}$

This table shows the pattern of meaning space usage averaged over 100 simulations with these parameters measured at 50 generations:

features values	1 256	2 16	3 6	4 4	5 3	6 3	7 3	8 2
average expressivity	0	0	0	.11	.29	.15	.03	.45

Despite being identical initially, agents end up using different systems of meaning for expressing objects in the environment in each simulation. In some runs, such as in figure 6, multiple meaning spaces remain partially expressive and stable. This means that agents may have different ways of expressing the same object. Real languages have different ways of carving up the world, and real speakers have different ways of expressing the same message. This simulation demonstrates a mechanism by which this can be acquired and can evolve culturally.

Are there any generalisations that can be made about the particular linguistic systems that emerge through this evolutionary process? A clear answer to this requires further research, but it may be that the meaning space adapts to structure in the environment. In the current model, the pairing between objects and points in meaning spaces is initialised randomly with uniform probability. A future version of the model will allow the experimenter to populate the environment with objects with non-uniform distribution in meaning space.

4.4 Inexpressive meaning spaces and the role of context

In this model, there is a many-to-one mapping from objects in the environment onto meanings in any one meaning space. This means that the simulation can be set up in such a way that agents can produce expressions that are hugely ambiguous. Conceivably, a meaning space could be available that mapped all the objects in the environment onto one point. We can think of an agent using such a meaning space as expressing every object as "thing".

What happens in the iterated learning model when these "inexpressive" meaning spaces are included? An experiment was run with the following parameters:

$$\mathrm{I} = 0.1,\ \mathrm{N} = 100,\ \mathrm{R} = 50,$$
$$\mathrm{C} = \{\mathrm{M}_{\langle 1,256\rangle}, \mathrm{M}_{\langle 8,2\rangle}, \mathrm{M}_{\langle 2,2\rangle}\}$$

In this situation, the agents end up expressing all 100 of the objects in the environment using the two-by-two meaning space. To put it another way, they use two word sentences with a vocabulary of four words. This kind of language is very stable since it requires very little data to learn.

This seems a rather implausible result. In reality, language is used to communicate rather than merely label objects. To simplify somewhat, in a particular situation, a speaker may attempt to draw a hearer's attention towards one of a range of possible objects in the current context.[5] If all the objects in the context map to the same meaning in the language, then no expression could be possible that would successfully direct the hearer's attention. Only if the context size was minimised could an inexpressive meaning space hope to discriminate the intended object from the others, but in the limit this essentially renders communication irrelevant. If there is only one possible object to talk about, then the hearer will already know what it is.

Contexts can be added to the simulation model relatively easily. Speakers are given a target object and a number of other objects that form the context. When choosing a meaning space to use to convey the target, speakers will reject meanings that fail to discriminate the target from one or more of the objects in the context.

Repeating the previous simulation with a context of 5 objects leads to the domination of the expressive eight-by-two meaning space over the inexpressive two-by-two one. This result demonstrates once again how iterated learning can result in language adapting over a cultural timescale to the particular constraints placed on its transmission.

5 Conclusions

In this paper I have shown how previous models of iterated learning which used monolithic meaning spaces can be extended to deal with a more flexible notion of meaning. By allowing agents choice over the semantics of linguistic expressions, we can see how meanings as well as signals evolve culturally.

This extension has allowed us to expand on earlier analyses of the relative stability of completely compositional versus completely holistic languages to look at the complete dynamics of a space of languages that are partially compositional. In addition, we can look at far more complex systems with ambiguity of meaning, varying degrees and types of

[5] Recall that "object" here is merely a term of convenience. We might wish to gloss this with "communicative intention".

compositionality and semantic structure, and examine how communicative contexts affect the way language is transmitted.

There is much work to be done in this area – I consider this model to be a preliminary investigation only. Many possible extensions of the model could be worth pursuing. For example, the results suggest a puzzle: why aren't all languages binary? The binary meaning spaces seem to be highly stable in the model, but nothing like this exists in natural language. What is needed is a more realistic treatment of semantics and also considerations of signal complexity. Natural language semantics does not take the form of fixed-length vectors, and there are plausible pressures to keep signals short.

Another interesting direction would be to combine this kind of idealised model with the mechanisms for collaborative meaning construction and grounding developed by those working with robotics models (e.g., Steels & Vogt, 1997; Steels, 1998; Vogt, 2003; Cangelosi, 2004). In this manner, we may begin to be able to relate abstract notions of expressivity, learnability and stability with the particular features of natural language semantics grounded in the real world and embodied in human agents.

The overarching conclusion of this line of work is that iterated learning is a surprisingly powerful adaptive system. The fact that language can only persist if it is repeatedly passed through a transmission bottleneck – the actual utterances that form the learning experience of children – has profound implications for its structure. This point has been made clear before in relation to the syntax of language. The model in this paper shows that the semantics of language are also likely to have been shaped by iterated learning.

References

Batali, J. Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M. and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases.* Cambridge: Cambridge University Press, 1998.

Brighton, H. Compositional Syntax from Cultural Transmission. *Artificial Life*, 8(1):25—54, 2002.

Brighton, H. *Simplicity as a Driving Force in Linguistic Evolution*. PhD thesis, Theoretical and Applied Linguistics, The University of Edinburgh, 2003.

Brighton, H. and Kirby, S. The Survival of the Smallest: Stability Conditions for the Cultural Evolution of Compositional Language. In J. Kelemen and P. Sosík, editors, *ECAL01*, pages 592--601. Springer-Verlag, 2001.

Cangelosi, A. The sensorimotor bases of linguistic structure: Experiments with grounded adaptive agents. In S. Schaal et al., editor, *SAB04*, pages 487--496. Los Angeles: Cambridge MA, MIT Press, 2004.

Chomsky, N. *The Minimalist Program.* Cambridge, MA: MIT Press, 1995.

Hurford, J. Expression/induction models of language evolution: dimensions and issues. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge University Press, 2002.

Kirby, S. Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303--323. Cambridge University Press, 2000.

Kirby, S. and Hurford, J. The Emergence of Linguistic Structure: An overview of the Iterated Learning Model. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pages 121--148. London: Springer Verlag, 2002.

Smith, A. D. M. Intelligent Meaning Creation in a Clumpy World Helps Communication. *Artificial Life*, 9(2):559—574, 2003.

Smith, K. *The Transmission of Language: models of biological and cultural evolution.* PhD thesis, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, 2003.

Steels, L. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1-2):133—156, 1998.

Steels, L. and Vogt, P. Grounding adaptive language games in robotic agents. In I. Harvey and P. Husbands, editors, *ECAL97.* Cambridge, MA: MIT Press, 1997.

Tonkes, B. *On the Origins of Linguistic Structure: Computational models of the evolution of language*. PhD thesis, School of Information Technology and Electrical Engineering, University of Queensland, Australia, 2001.

Vogt, P. Iterated Learning and Grounding: From Holistic to Compositional Languages. In Simon Kirby, editor, *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI.* Vienna, 2003.

Wray, A. Protolanguage as a holistic system for social interaction. *Language and Communication*, 18(1):47—67, 1998.

Zuidema, W. How the poverty of the stimulus solves the poverty of the stimulus. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02).* Cambridge, MA: MIT Press., 2003

Entropy Indicators for Investigating Early Language Processes

Caroline Lyon*, Chrystopher L. Nehaniv*, and Bob Dickerson*
*School of Computer Science
University of Hertfordshire
College Lane
Hatfield, Hertfordshire AL10 9AB
United Kingdom
C.M.Lyon@herts.ac.uk, C.L.Nehaniv@herts.ac.uk, R.G.Dickerson@herts.ac.uk

Abstract

We examine evidence for the hypothesis that language could have passed through a stage when words were combined in structured linear segments and these linear segments could later have become the building blocks for a full hierarchical grammar. Experiments were carried out on the British National Corpus, consisting of about 100 million words of text from different domains and transcribed speech. This work extends and supports the results of our previous work based on a smaller corpus reported previously. Measuring the entropy of the texts we find that entropy declines as words are taken in groups of 2, 3 and 4, indicating that it is easier to decode words taken in short sequences rather than individually. Entropy further declines when punctuation is represented, showing that appropriate segmentation captures some of the language structure. Further support for the hypothesis that local sequential processing underlies the production and perception of speech comes from neurobiological evidence. The observation that homophones are apparently ubiquitous and used without confusion also suggests that language processing may be largely based on local context.

1 Introduction

Hypotheses on the evolution of language can sometimes be supported, or undermined, by an investigation into underlying characteristics of present day language. Information theory provides some effective tools for carrying out such investigations, and is employed here as a tool for examining the hypothesis that the underpinnings of modern human language may lie in sequential processing phenomena, (though we also find that simple observations of every day speech can also be illuminating.)

1.1 Overview of the investigations

The core of the work described in this paper is an investigation into the statistical characteristics of spoken and written language which can help explain why language was likely to evolve with a certain structure. We take a large corpus of written text and transcribed speech and see whether the efficiency of encoding and decoding the stream of language is improved by processing a short sequence of words rather than individual words. To do this we measure the entropy of the word sequence, comparing values when we take single words, pairs, triples and quads. A decline in entropy indicates an increase in predictability, facilitating an improvement in decoding efficiency.

We also measure the entropy with and without punctuation, to see whether communication is more efficient if the stream of words is broken into segments that usually correspond to syntactic components. Our experiments (reported below) show that entropy does indeed decline as word sequences up to length three are processed, and thus supports the hypothesis that local sequential processing underpins communication through language. Entropy also declines further with the inclusion of punctuation. As there is a strong correlation between punctuation and prosodic markers in speech (Fang and Huckvale (1996); Taylor and Black (1998)) this decline indicates that there is an advantage in taking language in the segments that prosodic markers provide, since it is then easier to decode.

This suggests that there could be an intermediate stage in the development of a full hierarchical grammar. Processing a linear stream of words that is appropriately segmented is more efficient for the decoder than taking unsegmented, continuous strings of words. Such segments can then be the components of a hierarchical grammar.

Experiments have been carried out with the British National Corpus, BNC, about 100 million words of text and transcribed speech from many different domains (BNC).

1.2 Related work

We point to recent work on the "small world" phenomenon that investigates possible universal patterns of organization in complex systems (i Cancho and Sole (2001)). This effect, which is evident in natural language, picks up on the dominance of local dependencies, and research is going on into how robust complex systems can emerge, Section 5.

We also draw attention to other work that supports our hypotheses: neurobiological, computer modelling, and simple observation of everyday speech.

2 Background to this work

2.1 Co-operative communication

A number of scenarios have been used to introduce hypotheses on the evolution of language, and methods of communication between different animal species in different situations have been studied extensively. This has included a range of possibilities such as "gossip, deceit, alliance building, or other social purposes" (Bickerton (2002)). The work described here is based on those scenarios where producers and receivers are co-operating, sharing information. In the past little work in behavioural ecology had been done to make systematic comparisons of co-operative and non-cooperative signals (Krebs and Davies (1993)). A typical scenario for co-operative communication would be in group hunting or fishing situations, where deceit would be counter-productive. Even with manipulative communication a degree of co-operation is required to enable understanding. We look at modes of communication that are most efficient for producers and receivers. To investigate this we take a large corpus of spoken and written language and apply an analytic tool from information theory, the entropy measure, to help determine which possible characteristics of communication can make it more or less efficient.

2.2 Entropy indicators

The original concept of entropy was introduced by Shannon (1993)[1951]. Informally, it is related to predictability: the lower the entropy the better the predictability of a sequence of symbols. Shannon showed that the entropy of a sequence of letters declined as more information about adjacent letters is taken into account; it is easier to predict a letter if the previous ones are known. Entropy is represented as H, and we measure

- H_0 : entropy with no statistical information, symbols equi-probable.
- H_1 : entropy from information on the probability of single symbols occurring.
- H_2 : entropy from information on the probability of 2 symbols occurring consecutively.
- H_n : entropy from information on the probability of n symbols occurring consecutively.

More precisely, H_n measures the uncertainty of a symbol, conditional on its $n-1$ predecessors. (For $n > 0$, this is called the conditional entropy.)

For an introductory explanation of the concept of entropy, see (Lyon et al., 2003, page 170). The derivation of the formula for calculating entropy is in Appendix B. For many years Automated Speech Recognition developers have used entropy metrics to measure performance (Jelinek (1990)).

2.3 Using real language

A significant amount of language analysis in this field has not been done with real language. Well known examples include Elman's experiments with recurrent nets (Elman (1991)), which use a 23 word vocabulary: 12 verbs 10 nouns and a relative pronoun. Sentences like *boy sees boy* are considered grammatical, because there is number agreement between the subject and verb, though this sentence would be considered ungrammatical in real language with determiners missing. Elman himself is careful to say that this language is artificial, but this is not the case with many of his followers, who claim it is is a subset of natural language.

In fact many, sometimes most, of the words most people utter are function words. Though in any model we have to abstract out the features we consider most significant, we suggest that the common focus on content words introduces distortions. For example, to jump from words to syntactic combinations of nouns and verbs without considering the intermediate stage of phrase development leads to unrealistic conclusions. In our work we need to take language as it is.

3 The British National Corpus

Other recent work in this field has been done on a comparatively small corpus of 26,000 words of transcribed speech, annotated with prosodic markers (Lyon et al. (2003, 2004)). However, using the large BNC corpus enables us to confirm those results, and extend them.

The BNC corpus is composed of a representative collection of English texts; about 10% of the total is transcribed speech. As we want to investigate the processing of running language, headlines, titles, captions and lists are excluded from our experiments. Then adding in punctuation marks leads to a corpus of about 107 million symbols.

In order to carry out an analysis on strings of words it is necessary to reduce an unlimited number of words to a smaller set of symbols, and so words are mapped onto parts of speech tags. As well as making the project computationally feasible this approach is justified by evidence that implicit allocation of parts of speech occurs very early in language acquisition by infants, even before lexical access to word meanings (Morgan et al. (1996)).

The BNC corpus has been tagged, with a tagset of 57 parts of speech and 4 punctuation markers. We have mapped these tags onto our own tagset of 32 classes, of which one class represents any punctuation mark (Appendix A). Tag sets can vary in size but our underlying aim is to group together words that function in a similar way, have similar neighbours. Thus, for example, lexical verbs can usually have the same type of predecessors and successors whether they are in the present or past tense:

We like swimming / We liked swimming

so in our tagset they are in one class. This maintains a good degree of discriminability while moving to a smaller, fairly natural tagset. Moreover, another reason for mapping the BNC tagset onto our smaller set is that the entropy measures are more pronounced for the smaller set, while a larger tagset would require even larger corpora to avoid undersampling errors in entropy estimates.

4 Experiments

We have run the following experiments. First, we have processed the whole corpus of 107 million parts of speech tags, with punctuation, and found H_1, H_2, H_3, and H_4 as shown in Table 1. We also ran experiments over each of the 10 directories in which the corpus material is placed to see if there was much variation. In fact, variations between the directories is small: the results cluster round a central tendency shown by the measure for the whole corpus. An example is shown in Table 1.

We also process a comparable set of randomly generated numbers, in order to ensure that distortions do not occur because of undersampling. With 32 tags the number of possible sequences of length 5 are 33,554,432. If too small a sample is used the entropy appears lower than it should, since, e.g. not all the infrequent cases have occurred. A simple empirical test on sample size is through a random number sequence check. For a random sequence, the entropy should not decline as more of the information over preceding numbers is taken into account, since they are generated independently. Thus H for a sequence of random numbers in the range 0 to 31 should stay at 5.0. Sequences of random numbers are produced by the Unix random number generator. The results show that for the whole corpus we can be fully confident up to the H_4 figure, but H_5 should be treated with caution. For the 10 subdirectories, H_4 should be treated with caution, and H_5 is omitted.

Secondly, we process the whole BNC corpus, but omitting punctuation marks, as shown in Table 2. This time there will be 31 tags, as the punctuation symbol is omitted. The number of words is reduced, as punctuation marks are counted as words.

4.1 Analysis of results

The results in Table 1 show that entropy declines as processing is extended over the 1, then 2 and then 3 preceding consecutive parts of speech tags. There is a small further decline when 4 consecutive tags are taken. The results for 5 consecutive tags are not considered fully reliable, in view of the random sequence check for 107 million symbols.

Compare these results with those in Table 2. This time there is one less tag symbol, so we expect unpredictability to decrease compared to that for the corpus tagged with 32 symbols, and entropy to be less. This is what we find for H_0 and for H_1. However, as we take words 2, 3 and 4 at a time we find that entropy is slightly greater than in the first case. This

Corpus	H_0	H_1	H_2	H_3	H_4	H_5
107 million words + punctuation 32 tags	5.0	4.19	3.27	2.94	2.84	(2.75)
107 million random words 32 tags	5.0	5.0	5.0	5.0	5.0	4.8
10 million words, subdirectory F 32 tags	5.0	4.18	3.25	2.91	(2.79)	
10 million random words 32 tags	5.0	5.0	5.0	5.0	4.93	3.05

Table 1: Entropy measures for the BNC corpus, mapped onto 32 parts of speech tags. 3-grams, 4-grams and 5-grams that span a punctuation mark are omitted. Figures in brackets are to be treated with caution.

Corpus	H_0	H_1	H_2	H_3	H_4	H_5
94 million words, no punctuation 31 tags	4.95	4.16	3.29	3.14	3.07	(3.01)
94 million random words, 31 tags	4.95	4.95	4.95	4.95	4.95	4.72

Table 2: Entropy measures for the BNC corpus, mapped onto 31 parts of speech tags, omitting punctuation. The figure in brackets should be treated with caution.

indicates that punctuation captures some of the structure of language, allowing the next parts of speech tag to be be better predicted, and that by removing punctuation (corresponding to prosodic marking in speech) we increase the uncertainty. Paraphrasing Shannon we can say that a string of words between punctuation marks is a cohesive group with internal statistical influences, and consequently the n-grams within such phrases, clauses or sentences are more restricted than those which bridge punctuation ((Shannon, 1993, page 197)).

These results indicate that a stream of language is easier to decode if words are taken in short sequences rather than as individual items, and supports the hypothesis that local sequential processing underlies communication through language.

5 Other evidence for local processes

5.1 Computer modelling and the “small world” effect

In consider local processing, it is instructive to look at syntactic models based on dependency grammar and related concepts. Dependency grammar assumes that syntactic structure consists of lexical nodes (words) and binary relations (dependencies) linking them. Though these models are word based, phrase structure emerges. An online practical example is the Link Parser (Sleator et al. (2005)) where you can parse your own texts and see how the constituent tree emerges. Now, it is reported (i Cancho (2004)) that, in experiments in Czech, German and Romanian with a related system, about 70% of dependencies are between neighbouring words, 17% at a distance of 2.

This is one of the characteristics of the small world effect. A significant amount of syntactic knowledge is available from local information, even before our grammatical capability is enhanced by the addition of long range dependencies associated with phrase structure hierarchies.

From this one could also suggest that an intermediate stage in the development of a fully fledged grammar could have been based on local syntactic constraints.

Returning to another computer model, Elman's recurrent networks, we note that they could have a useful role to play in modelling short phrasal strings, but there are inherent obstacles to modelling longer dependencies (Bengio (1996); Hochreiter et al. (2001)).

5.2 Neurobiological evidence

Our hypothesis is also supported by the fact that primitive sequential processors in the basal ganglia play an essential role in language processing (Lieberman (2000, 2002)). The neural substrate that regulates motor control includes the control of articulatory acts, and this part of the brain seems to have extended its role to manage the sequencing of linguistic elements. An overview of the evidence that language and motor abilities are connected is given in a special edition of *Science* (Holden (2004)).

5.3 Simple observations of everyday speech

Any hypothesis on the evolution of language needs to explain why all languages seem to have homophones (Lyon et al. (2004)). In English some of the most frequently used words have more than one meaning such as *to / too / two*. Even young children seem able to disambiguate them without difficulty. In an agglutinative language such as Finnish they are rarely used by children, but occur in adult speech (Warren (2001)).

Their prevalence undermines the theories based on the assumptions that words in evolutionarily advanced language have a single meaning, that "the evolutionary optimum is reached if every word is associated with exactly one signal" ((Nowak et al., 1999, page 151)) and that there is a "loss of communicative capacity that arises if individual sounds are linked to more than one meaning" ((Nowak et al., 2002, page 613)). While such theories and models may appear to be logically attractive, they do not represent real language.

However, if we accept the hypothesis that local sequential processing underlies our language capability then there is not a problem accounting for the homophone phenomenon: homophones can be disambiguated by the local context.

6 Conclusion

When we look for clues to the evolution of language we can examine the state humans are in now and reason about how we could have arrived at the present position. This may take the form of brain studies, but it can also include the sort of analysis of language that we are doing. Chomsky once famously claimed that "One's ability to produce and recognize grammatical structures is not based on notions of statistical approximation and the like" (Chomsky (1957)). However, statistics can illuminate the way in which language processing has been carried out, and investigations on large corpora can now be done that were not possible a few decades back.

Our experiments suggest that utterances are processed in segments of a few words. We go on to hypothesize that these segments could be the elements out of which a hierarchical grammar is built.

References

T C Bell, J G Cleary, and I H Witten. *Text Compression*. Prentice Hall, 1990.

Yoshua Bengio. *Neural Networks for Speech and Sequence Recognition*. ITP, 1996.

Derek Bickerton. Foraging versus social intelligence in the evolution of protolanguage. In Alison Wray, editor, *The Transition to Language*, pages 207–225. OUP, 2002.

The British National Corpus. The BNC Consortium, http://www.hcu.ox.ac.uk/BNC.

N Chomsky. *Syntactic Structures*. The Hague: Mouton, 1957.

J L Elman. Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning*, pages 195–223, 1991.

Alex Chengyu Fang and Mark Huckvale. Synchronising syntax with speech signals. In V.Hazan et al., editor, *Speech, Hearing and Language*. University College London, 1996.

S Hochreiter, Y Bengio, P Frasconi, and J Schmidhuber. Gradient flow in recurrent nets: the difficuly of learning long term dependencies. In S C Kremer and J F Kolen, editors, *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press, 2001.

Constance Holden. The origin of speech. *Science*, 303:1316 –1319, 2004.

R Ferrer i Cancho. Patterns in syntactic dependency networks. *Physical Review E*, 69:051915, 2004.

Ramon Ferrer i Cancho and Ricard V. Sole. The small world of human language. *Proceedings of The Royal Society of London. Series B, Biological Sciences*, 268(1482):2261–2265, 2001.

F Jelinek. Self-organized language modeling for speech recognition. In A Waibel and K F Lee, editors, *Readings in Speech Recognition*, pages 450–503. Morgan Kaufmann, 1990. IBM T.J.Watson Research Centre.

J R Krebs and N B Davies. *An Introduction to Behavioural Ecology*. Blackwell, 1993.

P Lieberman. *Human Language and our Reptilian Brain*. Harvard University Press, 2000.

P Lieberman. On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology*, 2002.

C Lyon, B Dickerson, and C L Nehaniv. The segmentation of speech and its implications for the emergence of language structure. *Evolution of Communication*, 4, no.2:161–182, 2003.

C Lyon, C L Nehaniv, S Warren, and J Baillie. Evolutionary fitness, homophony and disambiguation through sequential processes. In *Proceedings of First International Workshop on Emergence and Evolution of Linguistic Communication, JSAI*, 2004.

J Morgan, R Shi, and P Allopenna. Perceptual bases of rudimentary grammatical categories. In J Morgan and K Demuth, editors, *Signal to Syntax*. Lawrence Erlbaum, 1996.

M A Nowak, N L Komaraova, and P Niyogi. Computational and evolutionary aspects of language. *Nature*, 417:611 – 617, 2002.

Martin A Nowak, Joshua B Plotkin, and David C Krakauer. The evolutionary language game. *J. Theoretical Biology*, 200:147–162, 1999.

C E Shannon. Prediction and Entropy of Printed English (1951). In N J A Sloane and Aaron D Wyner, editors, *Shannon: Collected Papers*. IEEE Press, 1993.

D Sleator, D Temperly, and J Lafferty. *Link Grammar*. Carnegie Mellon University, http://www.link.cs.cmu.edu/link/, 2005. Visited 17 Jan 2005.

P Taylor and A Black. Assigning phrase breaks from part-of-speech sequences. In *Proceedings of Eurospeech'97, Rhodes*, pages 995–998, 1998.

S Warren. *Phonological Acquisition and Ambient Language: a Corpus Based, Cross-Linguistic Exploration*. PhD thesis, University of Hertfordshire, UK, 2001.

Appendix A

The tagset of the British National Corpus is mapped onto our tagset. Each of the BNC tags is mapped onto an integer, as shown below, so that functionally similar tags are grouped together.

Tag **Code for our mapping**

AJ0 1
Adjective (general or positive) (e.g. *good, old, beautiful*)

AJC 1
Comparative adjective (e.g. *better, older*)

AJS 1
Superlative adjective (e.g. *best, oldest*)

AT0 2
Article (e.g. *the, a, an, no*)

AV0 3
General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. *often, well, longer* (adv.), *furthest*).

AVP 3
Adverb particle (e.g. *up, off, out*)

AVQ 3
Wh-adverb (e.g. *when, where, how, why, wherever*)

CJC 4
Coordinating conjunction (e.g. *and, or, but*)

CJS 4
Subordinating conjunction (e.g. *although, when*)

CJT 4
The subordinating conjunction *that*

CRD 2
Cardinal number (e.g. *one, 3, fifty-five, 3609*)

DPS 5
Possessive determiner-pronoun (e.g. *your, their, his*)

DT0 2
General determiner-pronoun: i.e. a determiner-pronoun which is not a DTQ or an AT0.

DTQ 2
Wh-determiner-pronoun (e.g. *which, what, whose, whichever*)

EX0 6
Existential *there*, i.e. *there* occurring in the *there is* ... or *there are* ... construction

ITJ 7
Interjection or other isolate (e.g. *oh, yes, mhm, wow*)

NN0 8
Common noun, neutral for number (e.g. *aircraft, data, committee*)

NN1 9
Singular common noun (e.g. *pencil, goose, time, revelation*)

NN2 10
Plural common noun (e.g. *pencils, geese, times, revelations*)

NP0 11
Proper noun (e.g. *London, Michael, Mars, IBM*)

ORD 1
Ordinal numeral (e.g. *first, sixth, 77th, last*) .

PNI 12
Indefinite pronoun (e.g. *none, everything, one* [as pronoun], *nobody*)

PNP 13
Personal pronoun (e.g. *I, you, them, ours*)

PNQ 14
Wh-pronoun (e.g. *who, whoever, whom*)

PNX 15
Reflexive pronoun (e.g. *myself, yourself, itself, ourselves*)

POS 16
The possessive or genitive marker *'s* or *'*

PRF 17
The preposition *of*

PRP 18
Preposition (except for *of*) (e.g. *about, at, in, on, on behalf of, with*)

PUL 0
Punctuation: left bracket - i.e. *(* or *[*

PUN 0
Punctuation: general separating mark - i.e. . , *!* , *:* *;* - or *?*

PUQ 0
Punctuation: quotation mark - i.e. *'* or *"*

PUR 0
Punctuation: right bracket - i.e. *)* or *]*

TO0 19
Infinitive marker *to*

UNC 7
Unclassified items which are not appropriately considered as items of the English lexicon.

VBB 20
The present tense forms of the verb BE, except for *is, 's*: i.e. *am, are, 'm, 're* and *be* [subjunctive or imperative]

VBD 20
The past tense forms of the verb BE: *was* and *were*

VBG 21
The *-ing* form of the verb BE: *being*

VBI 22
The infinitive form of the verb BE: *be*

VBN 23
The past participle form of the verb BE: *been*

VBZ 24
The *-s* form of the verb BE: *is, 's*

VDB 20
The finite base form of the verb DO: *do*

VDD 20
The past tense form of the verb DO: *did*

VDG 21
The *-ing* form of the verb DO: *doing*

VDI 22
The infinitive form of the verb DO: *do*

VDN 23
The past participle form of the verb DO: *done*

VDZ 24
The *-s* form of the verb DO: *does*, *'s*

VHB 20
The finite base form of the verb HAVE: *have, 've*

VHD 20
The past tense form of the verb HAVE: *had, 'd*

VHG 21
The *-ing* form of the verb HAVE: *having*

VHI 22
The infinitive form of the verb HAVE: *have*

VHN 23
The past participle form of the verb HAVE: *had*

VHZ 24
The *-s* form of the verb HAVE: *has, 's*

VM0 25
Modal auxiliary verb (e.g. *will, would, can, could, 'll, 'd*)

VVB 26
The finite base form of lexical verbs (e.g. *forget, send, live, return*) [Including the imperative and present subjunctive]

VVD 26
The past tense form of lexical verbs (e.g. *forgot, sent, lived, returned*)

VVG 27
The *-ing* form of lexical verbs (e.g. *forgetting, sending, living, returning*)

VVI 28
The infinitive form of lexical verbs (e.g. *forget, send, live, return*)

VVN 29
The past participle form of lexical verbs (e.g. *forgotten, sent, lived, returned*)

VVZ 30
The *-s* form of lexical verbs (e.g. *forgets, sends, lives, returns*)

XX0 31
The negative particle *not* or *n't*

ZZ0 7
Alphabetical symbols (e.g. *A, a, B, b, c, d*)

Appendix B

The derivation of the formula for calculating conditional entropy

This is derived from Shannon's work on the entropy of symbol sequences. He produced a series of approximations to the entropy H of written English, taking letters as symbols, which successively take more account of the statistics of the language.

H_0 represents the average number of bits required to determine a symbol with no statistical information. H_1 is calculated with information on single symbol frequencies; H_2 uses information on the probability of 2 symbols occurring together; H_n, called the n-gram entropy, measures the amount of entropy with information extending over n adjacent symbols. As n increases from 0 to 3, the n-gram entropy declines: the degree of predictability is increased as information from more adjacent symbols is taken into account. If $n-1$ symbols are known, H_n is the conditional entropy of the next symbol, and is defined as follows.

b_i is a block of $n-1$ symbols, j is an arbitrary symbol following b_i

$p(b_i, j)$ is the probability of the n-gram consisting of b_i followed by j

$p_{b_i}(j)$ is the conditional probability of symbol j after block b_i, that is $p(b_i, j) \div p(b_i)$

$$\begin{aligned} H_n &= -\sum_{i,j} p(b_i, j) * log_2\, p_{b_i}(j) \\ &= -\sum_{i,j} p(b_i, j) * log_2\, p(b_i, j) + \sum_{i,j} p(b_i, j) * log_2\, p(b_i) \\ &= -\sum_{i,j} p(b_i, j) * log_2\, p(b_i, j) + \sum_{i} p(b_i) * log_2\, p(b_i) \end{aligned}$$

since $\sum_{i,j} p(b_i, j) = \sum_i p(b_i)$.

N.B. This notation is derived from that used by Shannon. It differs from that used, for instance, by Bell et al. (1990).

Contextual semanticization of songstring syntax: a possible path to human language

Bjorn Merker
Department of Psychology
Uppsala University, Uppsala Sweden
gyr694c@tninet.se

Kazuo Okanoya
Brain Science Institute, Riken,
Fauclty of Letters, Chiba University, Japan
okanoya@cogsci.L.chiba-u.ac.jp

Abstract

The difficulties inherent in proceeding from the semantics of typically innate animal calls to lexical syntax encourage a search for alternative approaches to the animal origins of human language. One such is to start from the elaborate and structurally rich but un-semanticized sequences of song produced for display pursposes by some species of animals exhibiting vocal learning. Examples abound among birds, while whale and human song provide rare mammalian instances of the phenomenon. Humpback whales exhibit not only a rich repertoire of different learned songs, but these are shared among individuals of a group through copying. Such a shared repertoire of many distinct songstrings could become a vehicle for semantic communication if the use of such a repertoire became differentiated according to environmental, behavioral, or social context. That is, if different parts of the repertoire became selectively attached to specific contexts, different songs for different contexts, the songstrings would come to function as long and complex "song-names" for those contexts. If so, a further possibility arises: sub-sequences of these context-marking strings could be cross-matched to whatever contents, features or aspects that happen to be held in common by the contexts marked by the full strings (assumed to be not only structurally rich but also redundant). This supplies the logic of a stage-wise progression of parsing, cross-matching and contextual differentiation of increasingly specialized substring reference which we propose carries the potential of arriving at full lexical syntax, including the final stages of grammaticization described by historical linguists. This conception will be elaborated with the help of animal examples attesting to the biological plausibility of this scenario for the origins of human language.

1 Introduction

Human languages are quintessentially historical phenomena. Every known aspect of linguistic form and content is subject to change in historical time (Lehmann, 1995, p. 34). Many syntactic phenomena find their explanation in the historical processes that generated them, while apart from their history they can only be formalized (DeLancey, 1993). In fact, it has been proposed that the true universals of language are the modes of historical change that result in extant patterns of linguistic structure (Bybee, 2005).

If no aspect of language remains untouched by history, how far back does this process take us? Presumably all the way to the origin of language, since any stage in a historical process is preceded by earlier history. Yet at some point of the past there was no human language, and the question arises: out of what prior medium might human language have arisen in human prehistory?

In what follows we shall suggest that the transition to human, historical language had its point of departure in a prior historical medium featuring inter-generational transmission by non-genetic means, rich in syntactic complexity but originally devoid of semantics. Not only do such systems exist in non-human nature in the form of elaborate song-displays on the part of animals with vocal learning (Catchpole and Slater, 1995; Marler, 2000; Okanoya, 2002) but as we shall see, the semanticization of such a system to yield the essentials of language (Merker, in press) appears easier to achieve than the syntactification of a prior syntax-less semantic system. We first consider the problems encumbering the latter.

2 The self-blocking path of animal semantics

Animal call systems feature a pair-wise relation between call type and meaning, to the virtual exclusion of sequence combinatorics (syntax) as a carrier of meaning (Marler, 2000, p. 37). Calls are typically monosyllabic and carry one category of "meaning" per call type (Marler, 2004). Whether they express the varied emotional states of the animal calling or signal external social or environmental circumstances (as in alarm calls; Seyfarth et al., 1980) animal calls are meaningful signals to conspecifics within hearing. Because of this semantic aspect, animal calls have been by far the favorite model for an animal precursor of the words of human language, particularly since the documentation of "functional reference" in animal calls (Marler et al., 1992). Yet calls and words differ in fundamental ways.

Animal calls tend to be innately based, while human words are invariably learned; they are also composites in a way in which animal calls are not (Hurford, 2004). Moreover, human words function as elements of sentences, yet the tight coupling between sound gesture and emotional-motivational dynamics characteristic of animal calls prevents them from acting as carriers of meaning in a sequential combinatorial system.

The problem is this: in order to function as elements of a combinatorial system for conveying meaning, vocal gestures must possess a modicum of neutrality and independence with respect to the basic emotional-motivational forces animating an animal's behavior. Consider an animal whose repertoire includes two typical calls, one meaning "food", normally given when the animal finds food, the other meaning "fear", an alarm call normally given when the animal is frightened, as by a predator. Why not combine these calls to generate four meanings from the two, as follows: the alarm call followed by the food call signals "fear-food", i.e. "fear-with-food", and hence "poisonous food", while the food call followed by the alarm call would mean "food-fear", i.e. "food-with-fear", and hence "prey"?

To do so would require the animal to use a call inherently linked to its *own* state of fear (as evoked by a predator) to signal something the animal has no reason to fear (its prey), and also to use a call charged with positive valence (finding food) to signal an object with strong negative valence (poison). An inherent feature of calls – their strong unitary referential loading, on a typically innate and emotionally charged basis – prevents the animal from doing so. Thus the very traits which make calls so economical and efficient as components of an animal call system erect a functional barrier to the adoption of call combinations as a means to multiply meanings, and this blocks their development into a syntactic system. In searching for the path on which one single species in the entire history of life on earth came to supplement its call system with language we turn then from calls to that other use to which some animals put their voice, namely the learned vocal displays of animal song. Here a radically different situation obtains.

3 Syntax without semantics: The complexities of learned song

The song-bouts of many birds and a very few mammals abound in sequence complexity achieved by rearrangement of a finite set of song elements or phrases. Thus the sedge warbler varies the sequence of a repertoire of some 50 different song elements, producing song patterns which essentially never repeat (Catchpole, 1976). In the case of Bengalese finches the nature of their endlessly varying non-deterministic song sequences has been formally worked out: it conforms to the output of a finite state grammar (Okanoya, 2002), and may even require a context free grammar to be fully described (Okanoya, unpublished observations). In the case of the humpback whale, one of the few mammals with complex learned song, the patterns that result from song learning are also shared. Individual males introduce innovations in their song, and the novelty spreads gradually through the group by copying (Payne, 2000; Okanoya, unpublished observations). The resulting repertoire is a largely shared and slowly changing set of complex and syntactically structured vocal strings specific for a given group and time, whose transmission across the seasons and generations amounts to a true song culture.

These and other feats of animal song, such as the 1800 distinct melodies making up the repertoire of the brown thrasher (Kroodsma & Parker, 1977), all depend on the singular biological trait of vocal learning. This highly specialized mechanism equips an animal with the readiness and ability to learn to shape its vocal output to match and duplicate patterns of auditory models received through the sense of hearing and incorporating arbitrary pattern characteristics (see Ziegler and Marler, 2004). Humans have this capacity while chimpanzees do not (Snowdon and Elowson, 1992; Janik & Slater, 1997), and we depend on it for every word we know how to pronounce and for every song we sing. If this capacity seems unremarkable to us, it is only because we have it. Yet in phylogenetic perspective it is exceedingly rare. No dog has ever been heard to imitate a cat or a crow, while mynah birds imitate all three. Mammals excel in their capacity to learn, yet *vocal* learning is a rarity among them (see review by Janik and Slater, 1997). Beyond a few mammals such as humans, whales and bats, it is birds that supply the striking examples of the ability to learn vocal patterns from auditory models.

The learned nature of these song displays, the existence of a practice phase resembling babbling ("sub-song"), the open ended pattern generativity in the singing of some species, and the cultural transmission of song traditions, specific to subpopulations (often in the form of dialects, Nottebohm, 1970) has repeatedly led students of bird song to note its relevance – in both ontogenetic and evolutionary terms – to fundamental issues in the origins of human language (Darwin, 1871; Marler, 1970; Nottebohm, 1975; Doupe and Kuhl 1999; Okanoya, 2002; Wilbrecht and Nottebohm 2003; Jarvis, 2004). What is entirely missing in this regard, of course, is any hint of semantic differentiation in animal song: in sharp contrast to the situation regarding animal calls, there is no evidence that animal song performs semantic functions in the sense of carrying functional reference (Marler, 2000).

The pattern variety of song could in principle convey vast amounts of specific information, yet nothing indicates that it is put to such use in nature. Being a product of sexual selection, song essentially serves to impress, be it rivals or potential mates (Catchpole and Slater, 1995). Song syntax thus offers an unexploited resource for use as an informational medium, provided it could be semanticized. How might this come about?

4 Songstring semanticization by contextual repertoire differentiation: moving song from aesthetics to language

Let humpback whales, instead of discarding old portions of their repertoire as new additions are made through innovation, attach some portions of their repertoire to one context and other portions to other context, by preferentially using certain song-strings in one context but not in another. There is precedent for such a process, if only incipiently so, in the singing of the wood warbler. Many of them do not use the same repertoire of song types indiscriminately for mate attraction and for fending off rival males, but tend to use some song types more in one context than in the other, perhaps reflecting differences in motivational state (Kroodsma, 1988). In the setting of the shared repertoire possessed by a group of humpback whales, the extension of such repertoire differentiation to multiple behavioral and environmental contexts would eventually make small sets of song-strings or single song-strings into "markers" or long "song-names" for those contexts or circumstances to which they were attached through preferential use.

Whether we use whales or birds as model, these strings, now preferentially attached to certain contexts, are composed of a finite set of unitary elements in long sequences which are not only highly varied, but redundant as well (Okanoya, 2002). This means that despite the situation-specificity of each songstring, these strings would contain substrings and partial sequences that were shared across strings – and therefore situations – in initially unpredictable patterns. This in turn would provide the opportunity to start collating subsequences of matching substrings with matching situational characteristics in a process which thereby would start incorporating aspects of categories and relations - such as kinds of actions, agencies, objects and attributes - cutting across situations. Initially such cross-fitting between songstrings and the world could proceed quite freely, because the setting was that of frivolous song rather than of serious meaning. Poor fits would have occasioned no disadvantage compared to the unsemanticized prior state, yet the more apposite the cross-matches, the more would this inchoate meaning-aspect start constraining the process of collation and cross-fitting.

Imagine, if you will, a nascent social game whose aim it was to discover similarities across disparate contexts and marking these by distinctive song fragments. Pursuing such a game, perhaps as a means to impress in the setting of sexual selection, one could do worse than to adopt the trick of the winter wren: to cut up the received songs into sections for rearrangement (see Kroodsma & Momose, 1991; Marler, 1991). Such "parsing" and reshuffling would, over time, allow cognitive categories capturing significant aspects of the workings of the world to be matched by corresponding substructures in the shared song repertoire. Proceeding to ever finer levels of correspondance through the process of subdivision and rearrangement, sometimes with deliberate invention – and social imposition? – of a string-usage convention, this might eventually yield something resembling single words in the lexicon of human language.

How much further than that would this process have to be taken in order to qualify as human language? As far as the natural history of the process is concerned, we are actually done, because the state of song-string semanticization we have just outlined corresponds quite closely to the historical starting point of the various processes of grammaticization studied by historical linguists. As one of them has expressed it: "Thus we assume that grammaticalization starts from a free collocation of potentially uninflected lexical words in discourse" (Lehmann, 1995, p. 12). Such "free collocations of uninflected lexical items in discourse" is what the above process of song-string parsing and cross-matching to ever more differentiated contexts would deliver, and this in the setting of a shared, now semanticized, repertoire bequeathed through vocal learning to each new generation in a cumulative, historical process of cultural transmission.

The above, we suggest, constitutes a possible path to the biological origin of human language. Needless to say it is only the abstract principle of such a path that we have sketched, and not its implementation in terms of its cognitive prerequisites and logistical details, nor in terms of the sequence of the human fossil and archaeological record, and the neurological and behavioral developments to which that record bears witness. Such issues belong to the future. For now we rest content with a conjecture which at least has the merit of steering clear of biological magic.

Acknowledgements

The writing of this contribution was supported by a grant from the Bank of Sweden Tercentenary Foundation to Bjorn Merker.

References

Bybee, Joan (2004). Mechanisms of change as universals of language. Article to appear in Spanish, English version available for download at http://www.unm.edu/~jbybee/

Catchpole, C. K. (1976). Temporal and sequential organization of song in the sedge warbler (Acrocephalus schoenobaenus). *Behaviour, 59,* 226-246.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge, UK: Cambridge University Press.

Darwin C (1871). *The descent of man and selection in relation to sex*. D. Appleton & Company, New York.

DeLancey, Scott (1993). Grammaticalization and Linguistic Theory. In J. Gomez de Garcia and D. Rood (Eds.), Proceedings of the 1993 Mid-America Linguistics Conference, Department of Linguistics, University of Colorado, pp. 1-22.

Hurford JR (2004). Human uniqueness, learned symbols and recursive thought. *European Review, **12**,* 551-565.

Janik, V. M. and Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior, 26,* 59-99.

Jarvis ED (2004). Learned birdsong and the neurobiology of human language. In HP Ziegler and P Marler, eds. Behavioral Neurobiology of Birdsong, pp. 749–777. *Annals of the New York Academy of Sciences, **1016***.

Kroodsma, D. E. (1988). Song types and their use: Developmental flexibility of the male Blue-winged Warbler. *Ethology, 79,* 235-247.

Kroodsma, D. E. and Momose, H. (1991). Songs of the Japanese population of the winter wren (Troglodytes troglodytes). *Condor, 93,* 424-432.

Kroodsma, D. E. & Parker, L. D. (1977). Vocal virtuosity in the brown thrasher. *Auk, 94,* 783-785.

Lehmann, Christian (1995): Thoughts on Grammaticalization. Second, revised edition. München: LINCOM Europa.

Marler, P. (1970). Bird song and speech development: could there be parallels? *American Scientist, 58,* 669-673.

Marler, P. (1991). Song-learning behavior: The interface with neuroethology. *Trends in the Neurosciences, 14,* 199-206.

Marler, P. (2000). Origins of music and speech: Insights from animals. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The origins of music,* (pp. 31-48). Cambridge, MA: MIT Press.

Marler, P., Evans, C. S., and Hauser, M. D. (1992). Animal signals. Reference, motivation or both? In H. Papousek, U. Jürgens, and M. Papousek (Eds.) Nonverbal vocal communication: Comparative and developmental approaches (pp. 66-86). Cambridge, UL: Cambridge University Press.

Merker, B. (in press). From ape culture to ritual culture and language *via* vocal learning for song: The singular path to human uniqueness. In S. Malloch and C. Trevarthen (Eds.), Communicative Musicality. Oxford: Oxford University Press.

Nottebohm, F. (1970). Ontogeny of bird song. *Science, 167,* 950-956.

Nottebohm, F. (1975). A zoologist's view of some language phenomena, with particular emphasis on vocal learning. In E. H. Lenneberg, & E. Lenneberg (Eds.), *Foundations of language development*, pp. 61-103. New York: Academic Press.

Okanoya, K. (2002). Sexual Display as a Syntactic Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection. In A. Wray (Ed.), *The transition to language*, pp. 44-64. Oxford: Oxford University Press.

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In N.L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music,* pp. 135-150. Cambridge, MA: The MIT Press.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour, 28,* 1070-1094.

Wilbrecht L and Nottebohm F (2003). Vocal learning in birds and humans. *Mental Retardation and Developmental Disabilities Research Reviews,* ***9****,* 135–148.

Zeigler, H. P. & Marler, P. (Eds.) (2004). Behavioral Neurobiology of Birdsong. *Annals of the New York Academy of Sciences, 1016.*

Language Change in Modified Language Dynamics Equation by Memoryless Learners

Makoto Nakamura[⋆] Takashi Hashimoto[†] Satoshi Tojo[⋆]
Japan Advanced Institute of Science and Technology
School of { [⋆]Information , [†]Knowledge } Science
Nomi, Ishikawa, 923-1292, Japan
{mnakamur, hash, tojo}@jaist.ac.jp

Abstract

Language change is considered as a transition of population among languages. The language dynamics equation represents such a transition of population. Our purpose in this paper is to develop a new formalism of language dynamics for a real situation of language contact. We assume a situation that memoryless learners are exposed to a number of languages. We show experimental results, in which contact with other language speakers during acquisition period deteriorates the learning accuracy and prevents the emergence of a dominant language. If we suppose a communicative language, when learners are frequently exposed to a variety of languages, the language earns relatively higher rate of population. We discuss the communicative language from the viewpoint of the language bioprogram hypothesis.

1 Introduction

In general, all human beings can learn any human language in the first language acquisition. One of the main functions of language use is to communicate with others. Therefore, it is easy to consider that the language learners come to obtain a language which they hear most in the community. In other words, the most preferable language in the community would eventually survive and become dominant in competition with other languages, depending on how much ratio of the people speak it. Accordingly, language change can be represented by a population dynamics, examples of which include an agent-based model of language acquisition proposed by Briscoe (2002) and a mathematical framework by Nowak et al. (2001), who elegantly presented an evolutionary dynamics of grammar acquisition in a differential equation, called the *language dynamics equation*.

Our purpose of this study is to develop a new formalism of language dynamics which deals with language contact between language learners and speakers, and then to investigate the relationship between the language contact and language change. Thus far, we have revised the model of Nowak et al. (2001) to be more realistic, in order to study the emergence of creole (DeGraff, 1999) in the context of population dynamics (Nakamura et al., 2003). For the purpose of modeling the process of creolization, we claimed that children during language acquisition should contact not only with their parents but also with other language speakers. To meet this condition, we revised the transition rate between languages to be sensitive to the distribution of languages in the population at each generation. We introduced the *exposure rate* to determine the degree of influence from other languages during acquisition. Namely, focusing on language learners, we have given a more precise environment of language acquisition than Nowak et al. (2001). In other words, introducing the exposure rate, we have regarded the model of Nowak et al. (2001) as a specific case of ours in language acquisition. Therefore, these revisions enable us to deal not only with the emergence of creole but also with other phenomena of language change.

In this paper, we aim at examining the behavior of our model in terms of language change. Komarova et al. (2001) adopted two kinds of language learners called *memoryless learners* and *batch learners*, comparing conditions of the two models for the emergence of a dominant language. In this paper, introducing a new transition probability for a memoryless learner exposed to a variety of languages, we compare the behavior of the dynamics with that of Komarova et al. (2001).

In Section 2, we propose a modified language dynamics equation and a new transition matrix of memoryless learning algorithm. We describe our experiments in Section 3. We discuss the experimental results in Section 4. Finally, we conclude this paper in Section 5.

2 Learning Accuracy of Memoryless Learners

2.1 Outline of the Language Dynamics Equation

We explain the outline of the language dynamics equation proposed by Nowak et al. (2001). In their model, given the principles in the universal grammar, the search space for candidate grammars is assumed to be finite, that is $\{G_1, \ldots, G_n\}$. The language dynamics equation is given by the following differential equations:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi x_i \quad (i = 1, \ldots, n), \tag{1}$$

where

x_i : the ratio of the population of G_i speakers, where $\sum_{j=1}^{n} x_j = 1$,

$Q = \{Q_{ij}\}$: the transition probability between grammars that a child of G_i speaker comes to acquire G_j,

f_i : fitness of G_i, which determines the number of children individuals reproduce, where $f_i = \sum_{j=1}^{n}(s_{ij} + s_{ji})x_j/2$,

$S = \{s_{ij}\}$: the similarity between languages, which denotes the probability that a G_i speaker utters a sentence consistent with G_j, and

ϕ : the average fitness or *grammatical coherence* of the population, where $\phi = \sum_i x_i f_i$.

The language dynamics equations are mainly composed by (i) the similarity between languages as the matrix $S = \{s_{ij}\}$ and (ii) the probability that children fail to acquire their parental languages as the matrix $Q = \{Q_{ij}\}$. The accuracy of language acquisition depends on the search space $\{G_1, \ldots, G_n\}$, the learning algorithm, and the number of input sentences, w, during language acquisition.

As a similarity matrix, in this paper, we mainly deal with such a special case that:

$$s_{ii} = 1, \quad s_{ij} = a, \quad i \neq j \ , \tag{2}$$

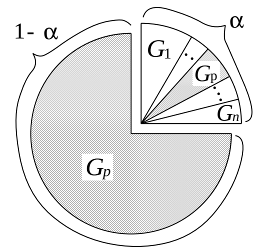


Figure 1: The exposure rate α

where a is a number between 0 and 1. In accordance, the transition probability comes to:

$$Q_{ii} = q, \quad Q_{ij} = \frac{1-q}{n-1}, \quad i \neq j \ , \tag{3}$$

where q is the probability of learning the correct grammar or the *learning accuracy* of grammar acquisition.

2.2 Modified Language Dynamics Equation

In a situation of language contact, a child may learn language not only from his parents but also from other language speakers who speak a different language from his parental one. In order to incorporate this possibility to language dynamics equation, we divide the language input into two categories; one is from his parents and the other is from other language speakers. We name the ratio of the latter an *exposure rate* α. This α is subdivided into the smaller ratios corresponding to the distribution of all language speakers. An example distribution of languages is shown in Fig. 1. The child of G_p speaker is exposed to G_p at the rate of the shaded part, that is $\alpha x_p + (1-\alpha)$, and the ratio of a non-parental language G_j comes to be αx_j.

Suppose that a child whose parents speak G_p hears sentences from the adult speakers depending on the exposure rate and on the distribution of population. If the child presumes G_j and hears a sentence, it is accepted with such a probability, U_{pj}, that:

$$U_{pj} = \alpha \sum_{k=1}^{n} s_{kj} x_k + (1-\alpha) s_{pj} \ . \tag{4}$$

For the special case where Eqn (2) is satisfied, it is transformed to:

$$U_{pj} = \begin{cases} 1 - \alpha(1-a)(1-x_j) & (p = j) \\ a + \alpha(1-a)x_j & (p \neq j) \end{cases} , \tag{5}$$

When a learning algorithm is expanded into the one which allows language learners to be exposed to a number of languages, the matrix $U = \{U_{ij}\}$ corresponds to $S = \{s_{ij}\}$ in terms of an acceptable probability of a sentence for a child. Then, the Q matrix depends on the U matrix and the U matrix on the population rate. Since the distribution of population changes in time, the Q matrix comes to include a time parameter t, that is, Q is redefined as $\overline{Q}(t) = \{\overline{Q}_{ij}(t)\}$. Thus, the new language dynamics equation is expressed by:

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} x_j(t) f_j(t) \overline{Q}_{ji}(t) - \phi(t) x_i(t) \quad (i = 1, \ldots, n). \tag{6}$$

We call it the *modified language dynamics equation*.

2.3 Memoryless Learning Algorithm

Komarova et al. (2001) argue two extreme learning algorithms called the batch learning algorithm and the memoryless learning algorithm (Niyogi, 1998), in which the former is considered as the most sophisticated algorithm within a range of reasonable possibilities, and the latter as the simplest mechanism. Because the memoryless learning algorithm is easy to be remodeled with our proposal, we will use it and compare the behavior of the dynamics with that of Komarova et al. (2001). In this section, we explain the learning accuracy of the memoryless learning algorithm derived from a Markov process.

The memoryless learning algorithm describes the interaction between a child learner and language speakers. Namely, the child hears sentences of a language. The learner starts presuming a grammar by randomly choosing one of the n grammars as an initial state. When the learner hears a sentence from the teacher, he tries to apply his temporary grammar to accept it. If the sentence is consistent with the learner's grammar, no action is taken; otherwise the learner changes his hypothesis about the grammar to the next one randomly picked up from the other grammars. This series of learning is repeated until the learner receives w sentences.

Komarova et al. (2001) supposed there is one teacher (the learner's parent), so that the learner hears only one language. In this case, the algorithm is presented by the following expressions. The initial probability distribution of the learner is uniform: $\boldsymbol{p}^{(0)} = (1/n, \ldots, 1/n)^T$, where A^T is the transposed matrix of A, i.e., each of the grammars has the same chance to be picked at the initial state. If the teacher's grammar is G_k and the child hears a sentence from the teacher, the transition process from G_i to G_j in the child's mind is expressed by a Markov process with such a transition matrix $M(k)$ that:

$$M(k)_{ij} = \begin{cases} s_{ki} & (i = j) \\ \dfrac{1 - s_{ki}}{n-1} & (i \neq j) \end{cases} . \tag{7}$$

After receiving w sentences, the child will acquire a grammar with a probability distribution $\boldsymbol{p}^{(w)}$. Therefore, the probability that a child of G_i speaker acquires G_j after w sentences is expressed by:

$$Q_{ij} = [(\boldsymbol{p}^{(0)})^T M(i)^w]_j \ . \tag{8}$$

The transition probability of the memoryless learning algorithm depends on the S matrix. For instance, if the condition of Eqn (2) is satisfied, the off-diagonal elements of the Q matrix are also equal to each other, and Eqn (3) holds. Therefore, $q = Q_{ii}$ $(i = 1, \ldots, n)$ is derived as follows:

$$q = 1 - \left(1 - \frac{1-a}{n-1}\right)^w \frac{n-1}{n} \ . \tag{9}$$

This is the learning accuracy of memoryless learner.

If once a memoryless learner achieves his parental grammar, he will never change his hypothesis. Suppose there exist only two grammars, then the memoryless learner has two states in a Markov process, that is, a state for the hypothesis of his parental grammar, G_{parent}, and a state for the other grammar, G_{other}. The transition probability between the states is expressed by a Markov matrix $M = \{m_{ij}\}$ such that:

$$M = \begin{pmatrix} 1 & 0 \\ 1-a & a \end{pmatrix} , \tag{10}$$

where

m_{11}: the probability that a child who correctly guesses his parental grammar maintains the same grammar,

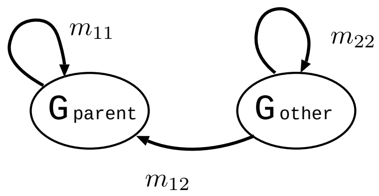
m_{12}: the probability that a child who correctly guesses his parental grammar changes his presumed grammar to another,

m_{21}: the probability that a child whose grammar is different from his parents' comes to presume his parental grammar, and

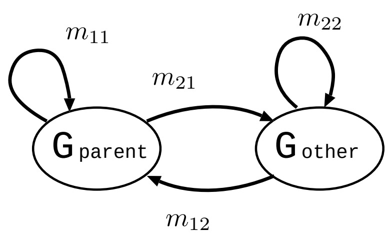
m_{22}: the probability that a child whose grammar is different from his parents' keeps the same grammar by accepting a sentence[1].

[1] If the memoryless learner is able to choose the refused grammar again with a uniform probability when he failed to accept the sentence, the Markov matrix is replaced by:

$$M = \begin{pmatrix} 1 & 0 \\ (1-a)/2 & a + (1-a)/2 \end{pmatrix} .$$



(a) A case a child hears sentences only from his parents



(b) A case a child hears sentences in a number of languages

Figure 2: Markov processes for the memoryless learning algorithm

Figure 2(a) shows a state transition diagram.

Komarova et al. (2001) have analyzed the language dynamics equation and deduced the following results:

- When the learning accuracy is high enough, most of the people use the same language, that is, there exists a dominant language. Otherwise, all languages appear at roughly similar frequencies.
- The learning accuracy is calculated from a learning algorithm. Receiving input sentences, a memoryless learner enhances his learning accuracy.

2.4 Memoryless Learners Exposed to a Number of Languages

We define a transition matrix, $\overline{Q}(t) = \{\overline{Q}_{ij}(t)\}$, of memoryless learners exposed to a number of languages during acquisition period. For a child whose parents speak G_p, the transition matrix of a Markov process is defined by:

$$M(p)_{ij} = \begin{cases} U_{pi} & (i = j) \\ \dfrac{1 - U_{pi}}{n-1} & (i \neq j) \end{cases} . \qquad (11)$$

The learning accuracy is derived by substituting Eqn (11) for Eqn (8). Because U_{ij} varies according to the distribution of population of each grammar, even in the special case where Eqn (2) is satisfied the learning accuracy of each grammar is different from each other. In other words, there are n values of the learning accuracy for each grammar. Expression (11) becomes equivalent to Eqn (7) at $\alpha = 0$. Thus, the transition probability with the exposure rate α is regarded as a natural extension of that of Komarova et al. (2001).

For a learner exposed to a variety of languages, the most important difference from a non-exposed learner is that even when the learner presumes his parental grammar G_p, a received sentence may not be accepted by the grammar with the probability $1-U_{pp}$. In this case he chooses one of the non-parental grammars randomly with a uniform probability. Thus, the memoryless learner is likely to refute his hypothesis even if once he acquired his parental grammar. In a two-grammars case, for example, the Markov matrix of this process is expressed by the following equation:

$$M(p) = \begin{pmatrix} U_{p1} & 1 - U_{p1} \\ 1 - U_{p2} & U_{p2} \end{pmatrix} . \qquad (12)$$

Figure 2(b) shows a state transition diagram of a memoryless learner exposed to a number of languages, which differs from Fig. 2(a) in that learners at a state G_p are possibly to move to another state.

In this section, we revised the memoryless learning algorithm in order to model a more real situation of language contact. In the next section, we examine how a memoryless learner is influenced by a variety of languages, and how a dominant language appears dependent on the initial conditions. Especially, we will look into the relationship between the exposure rate and the occurrence of a dominant language.

3 Experiments

In this section, we show that the behavior of our model with the memoryless learning algorithm depends on the exposure rate α. We set the number of grammars, $n = 10$, through the experiments. Firstly, comparing the dynamics of the model with that of Komarova et al. (2001), we examine how the exposure rate α works in our model. Secondly, we observe the behavior of dynamics, when there is a particular language in terms of the similarity.

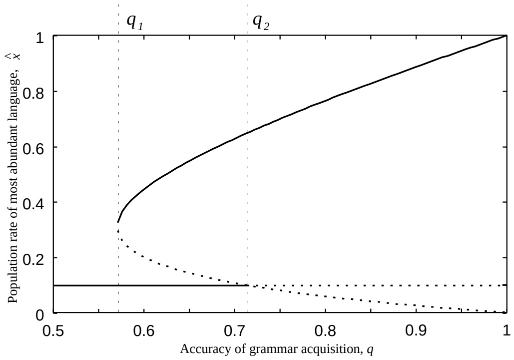


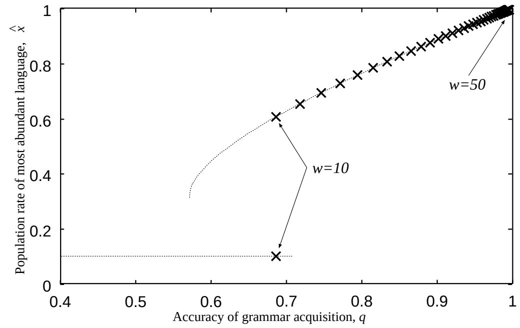
Figure 3: Analytical solutions of Eqn (1) which satisfies Eqn (2) and Eqn (3) ($n = 10, a = 0.1$)

3.1 Exposure and Learning Accuracy

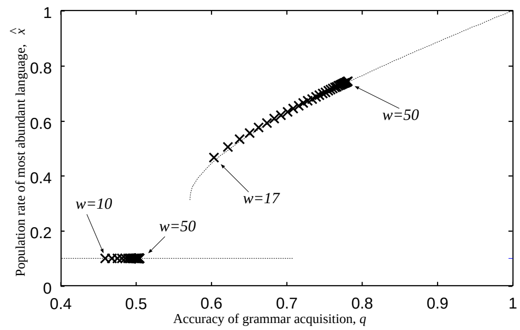
In this section, we observe the behavior of our model especially when Eqn (2) is satisfied. We compare the behavior of our model with analytical solutions of Komarova et al. (2001), and with the behavior of their model by memoryless learners, which is equivalent to that of our model at $\alpha = 0$.

Expression (1) substituted for Eqn (2) and Eqn (3) has analytically been solved by Komarova et al. (2001). The solutions of the model are derived by setting an arbitrary initial condition of the distribution of population, affected by the learning accuracy. Figure 3 shows the population rate of the most prevalent grammar in the community, $\hat{x}$, versus the learning accuracy, q, by which children correctly acquire the grammar of their parents, in case of $a = 0.1$. There are two types of solutions; one is that only one of the grammars earns a certain rate of population whereas the others are given the rest divided equally. Which of languages would be dominant depends on the initial condition. The other is that the solutions take the uniform distribution among grammars. Therefore, there are two thresholds, q_1 and q_2. When $q < q_1$, the population of each language would be uniform. When $q > q_2$, there would be one prevalent language in the community. Thus, q_1 is the necessary condition for the existence of the prevalent language and q_2 is the sufficient condition. When $q_1 < q < q_2$, the supremacy of one language depends on the initial distribution of population.

Here, we examined our model with memoryless learners at $\alpha = 0$, which is equivalent to that of Komarova et al. (2001). Because the learning accuracy, q, depends on the number of input sentence, w, the $q - \hat{x}$ relation is discretely represented by integer numbers of w. At $\alpha = 0$, the relation must identify that of the analytical solutions, depicted in Fig. 3.



(a) Solutions by memoryless learning ($\alpha = 0$)



(b) Solutions by memoryless learning ($\alpha = 0.12$)

Figure 4: The behavior of the model depending on the exposure rate α ($a = 0.1, w = 10, \ldots, 50$)

The result is shown in Fig. 4(a), in which the number of sentences, w, was given within the range from 10 to 50. In the figure, a cross ($\times$) denotes the $q - \hat{x}$ relation for a given w, and dotted lines are that of analytical solutions (copied from Fig. 3). As the result, we observed that the $q - \hat{x}$ relation of the model with memoryless learners exactly corresponds to that of the analytical solutions.

Next, we experimented different values of α in the memoryless learning by w. In our model, although the transition probability $\overline{Q}_{ij}(t)$ varies depending on the population rate at each generation, the value of $\overline{Q}_{ij}(t)$ becomes stable as the population rate approaches to the solution, and vice versa. Therefore, we can observe the $q - \hat{x}$ relation as well. We expected that because of the variable transition matrix $\overline{Q}(t)$, the $q - \hat{x}$ relation collapsed from that of the base model along with the increase of α. However, as is shown in Fig. 4(b) where $\alpha = 0.12$, the relation becomes the same as the one in Fig. 3. Instead, we can easily observe that the increase of α deterio-

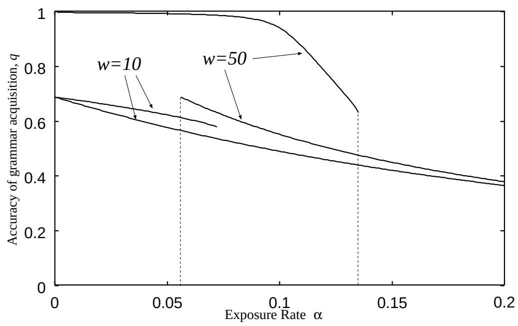


Figure 5: Exposure rate α versus learning accuracy q ($w = 10, 50$)

rates q in regard to w. Additionally, the solutions of q seem to be separated into two groups. We drew the graph with a several patterns of the initial distribution of population. As a result, some values of α seem to derive a bifurcation of q values which depend on the initial population distribution.

In order to observe the influence of α on q, we show $\alpha - q$ relation in Fig. 5, where two lines are represented for each of $w = 10$ and 50. The number of q values is determined according to α. At $w = 50$, when α is between the dashed lines in the figure, there exist two solutions of q which depend on the initial distribution of population. Accordingly, two solutions of $\hat{x}$ are derived at $\alpha = 0.12$ and $w = 50$, as shown in Fig. 4(b).

Although the $\alpha - q$ relation varies along with w, the learning accuracy, q, monotonously decreases depending on α, in common with any w. Therefore, the increase of α deteriorates q in regard to a common value of w.

In our model, q varies from generation to generation, while Komarova et al. (2001) gave a constant value to q fixed by a learning algorithm. We showed that q would be stable for given α and thus x also would be stable. Apparently $q - x$ relation is similar to that of Komarova et al. (2001). At this stage, we may well conclude that the increase of α would just decrease the accuracy of learning, and would not affect $q - x$ relation, when the algorithm is memoryless and the language similarity is uniform.

3.2 Communicative Language

In this section, we assume such a hypothetical language G_1, given G_2 and G_3, that is much similar to G_2 and G_3 than the rest. The S matrix is expressed by:

$$S = \begin{pmatrix} 1 & b & b & & \\ b & 1 & a & & a \\ b & a & 1 & & \\ & a & & \ddots & \\ & & & & 1 \end{pmatrix}, \qquad (13)$$

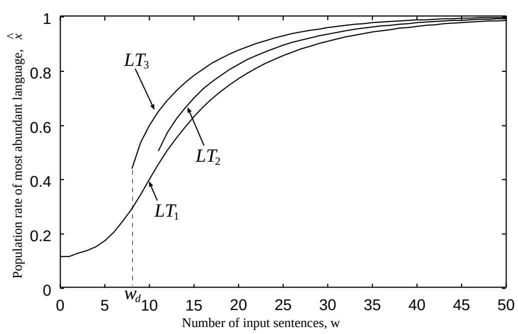
where $0 \leq a < b \leq 1$. We set $a = 0.1$ and $b = 0.5$ for the following experiments. Accordingly, languages are classified into three categories in terms of the similarity. For simplicity, we call them LT_1, LT_2 and LT_3, each of which includes the communicative language (G_1), the similar languages to G_1 (G_2 and G_3) and the others ($G_4 \ldots G_{10}$).

In order to observe how the exposure of children to a number of languages affects the most abundant language, we draw diagrams of the population rate of most prevalent language, $\hat{x}$, versus the number of input sentences, w, at particular points of α (see Fig. 6).

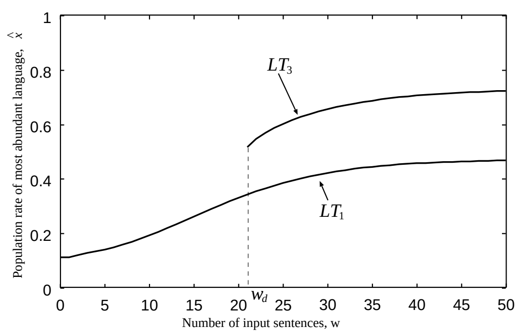
We start from $\alpha = 0$. Figure 6(a) shows that the greater the number of input sentences is, the higher the population rate of the most prevalent language exists in stable generations. The population rate of the most prevalent language depends on which of language types the language belongs to. Therefore, we can see three kinds of $w - \hat{x}$ relation in the figure, which correspond to the type of the language (LT_i). Note that in Fig. 6(a), $LT_1 < LT_2 < LT_3$. Komarova et al. (2001) explained the reason as follows; G_1 has a larger intersection with the rest of the languages than the rest of them. When this language becomes preferred, it stands out less than other languages would in its place, i.e., it corresponds to lower values of the population rate. When a language earns the most abundant population rate, the other languages share the rest, so that except for the most abundant language the rate of a language equals to another one which belongs to the same language type.

If w is smaller than a certain number, G_1 becomes the most abundant at any initial distribution of population. Otherwise, one of other languages might supersede G_1 depending on the initial condition. Here, we define a threshold w_d as the smallest number of input sentences in which a language other than G_1 could become the most prevalent language. When $\alpha = 0$, the threshold w_d is 8.

Figure 6(b) shows a diagram of $\hat{x}$ versus w at $\alpha = 0.12$. The threshold w_d is boosted to 21, and any of LT_2 does not earn the most abundant rate of population at $w < 50$. As was mentioned in Section 3.1, the increase of the exposure rate makes the learning accuracy low. For the memoryless learning algorithm, the learning accuracy, q, increases with



(a) Number of input sentences, w, versus population rate of most abundant language, $\hat{x}$ ($\alpha = 0$)



(b) Number of input sentences, w, versus population rate of most abundant language, $\hat{x}$ ($\alpha = 0.12$)

Figure 6: The behavior of the model with a communicative language

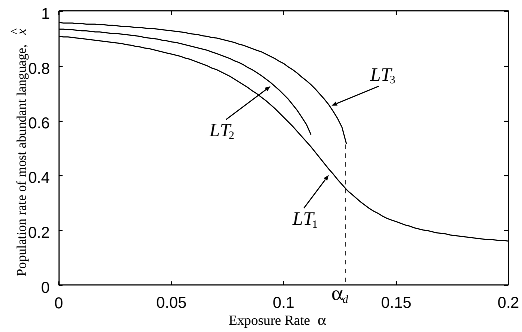


Figure 7: Influence of the exposure rate, α, on the population rate of most abundant language, $\hat{x}$ ($w = 30$)

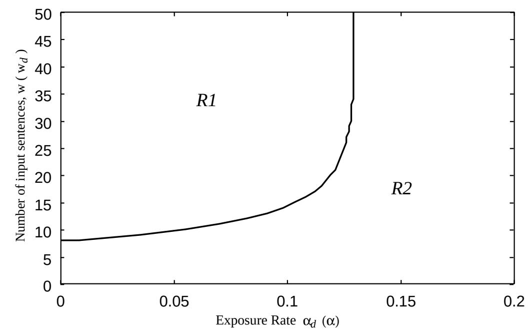


Figure 8: The relationship between two thresholds, α_d and w_d

the number of input sentences, w. The increase of w keeps the same quality of learning accuracy in response to α. Accordingly, w_d increases along with the exposure rate α.

We showed in Fig. 6 that the larger the exposure rate α was, the greater the threshold w_d was. It is expected that no matter how language learners are exposed to a number of languages, one of languages other than G_1 may stand out as long as the learners hear the proper quantity of language input. The quantity is w_d in Fig. 6. However, human beings have an acquisition period in which an appropriate grammar is estimated from their language input and it is limited in a finite time (Lenneberg, 1967). If the possible number of input sentences to be heard during acquisition period was settled in a specific value, then we could draw a diagram concerned with the influence of the exposure rate, α, on the population rate of most abundant language, $\hat{x}$. Figure 7 is an example of the diagram for $w = 30$.

We define α_d as the highest value of exposure rate at which one of languages other than G_1 could become the most abundant depending on the initial distribution. In case of $w = 30$, it was $\alpha_d \simeq 0.128$. It is easily conceivable that the greater the number of the input sentences is, the larger the threshold α_d is.

Thus far, we have observed the smallest number of input sentences for the appearance of the most abundant language other than G_1, that is w_d, at particular values of α. On the other hand, we saw the highest value of the exposure rate for the appearance of the most abundant language other than G_1, that is α_d, at a particular number of the input sentence. These two values have a functional relationship as shown in Fig. 8. This figure represents conditions of w and α on the appearance of the most abundant language other than G_1. The necessary number of input sentences rapidly increases along with the exposure rate. Learners need to receive 222 sentences at $\alpha = 0.13$, while 34 sentences at $\alpha = 0.129$. Although the $\alpha - w$ relation depends on the S matrix, the figure of the curve is expected to be basically kept at arbitrary distribution of elements in the S matrix.

4 Discussion

4.1 Possibility of Dominant Language

Figure 8 can be recognized as a boundary between the following two regions:

R1: one of languages other than the communicative one may become predominant.

R2: the communicative language obtains a certain rate of population for any initial conditions.

Language learners growing under the condition of $R1$ hear enough language input to acquire their parental languages with high learning accuracy. Although one of the languages may predominate in the community, which of languages becomes predominant depends on the initial distribution of population. Some of them are regarded as a dominant language. In most cases, the most populous language at the initial state tends to take the supremacy.

In the area of $R2$, the most populous language comes nothing but G_1, although it is hard to be regarded as a dominant language because of smaller population rate. Even if no one spoke G_1 at the initial state, G_1 eventually comes to be the most abundant language. Therefore, the change of the predominant language is easy to occur.

It seems that the condition of $R1$ is hardly satisfied for w when α is larger than approximately 0.13. This result suggests that any dominant language never appear as long as language learners are frequently exposed to a variety of languages.

4.2 Communicative Language and Bioprogram Hypothesis

In Section 3.2, we assumed that there is a communicative language, which is more similar to particular two languages than the others, that is G_1. Let us consider what the language corresponds to in the real world. We dare say that it is considered as a language that Bickerton (1984) supposed in the *Language Bioprogram Hypothesis*. Kegl et al. (1999) briefly outline the features of the hypothesis as follows:

> Bickerton (1984) proposed the Language Bioprogram Hypothesis. This hypothesis claims that a child exposed to nonoptimal or insufficient language input, such as a pidgin, will fall back on an innate language capacity to flesh out the acquisition process, subsequently creating a creole. This is argued to account for the striking similarities among creoles throughout the world.
>
> Kegl et al. (1999)

The communicative language has something in common with the bioprogrammed language in terms of the condition of existence; it appears when learners are frequently exposed to other languages so that any dominant language does not appear, or when they are not given sufficient language input. Therefore, if no one spoke the communicative language at the initial state, it would emerge as a creole.

If we recognize the communicative language to be consistent with the language bioprogram hypothesis, the bioprogrammed language is more communicative with pre-existing languages than the others. However, we cannot examine whether the creole is more similar to some particular languages or not. To ensure our hypothesis here, we need to embed linguistic features into the equation.

5 Conclusion

Contact of different language groups has been considered as one of main factors in language change. We modeled the contact by introducing the *exposure rate* to the language dynamics equation proposed by Nowak et al. (2001). The exposure rate is the rate of influence of languages other than the parental one on language acquisition. We assess the accuracy of parental language acquisition in the memoryless learning algorithm. The exposure to other languages made it possible that the language learner doubted his hypothetical grammar even though he once acquired his parental grammar. We expressed the acquisition process in a Markov matrix, and then revised a new transition probability that changes in accordance with the distribution of population, which is a different feature from Nowak et al. (2001). In addition, each grammar has a different learning accuracy even in the completely symmetrical similarity matrix of Eqn (2).

As the experimental result showed, the emergence of a dominant language depends not only on the similarities between languages but also on the ratio of contact of multiple languages.

We compared our result with Komarova et al. (2001) in Section 3.1. First, in case the similarity was uniform, we found that the introduction of the exposure rate α only deteriorated the accuracy of the target language acquisition; even though the population ratio versus the learning accuracy was the same, the introduction of α delayed the learning process. For memoryless learners, the failure of communication after achieving their parental grammars is fatal to the acquisition of a correct grammar. Therefore, when children are only exposed to other languages a little, a dominant language disappears. On the contrary, we expect that batch learners are robuster in terms of the noise. Our next target is to show the similar phenomena in the batch learning algorithm.

In the next experiment, we assumed that there is a most communicative language among the multiple language communities. The result suggests the following matters; If language learners hear enough language input to estimate their parental languages, one of languages other than the communicative language would be dominant. However, when language learners are frequently exposed to a variety of languages, the communicative language earns a certain rate of population regardless of the number of input sentences. The characteristic behaviors suggest that a bioprogrammed language hypothesized by Bickerton (1984). The experimental result shown in Fig. 8 suggests that creole will emerge when language learners are exposed to a variety of languages at a certain rate.

References

D. Bickerton. The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7(2):173–222, 1984.

E. J. Briscoe. Grammatical acquisition and linguistic selection. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 9. Cambridge University Press, 2002.

M. DeGraff, editor. *Language Creation and Language Change*. The MIT Press, Cambridge, MA, 1999.

J. Kegl, A. Senghas, and M. Coppola. Creation through contact: Sign language emergence and sign language change in nicaragua. In M. DeGraff, editor, *Language Creation and Language Change*, chapter 1. The MIT Press, Cambridge, MA, 1999.

N. L. Komarova, P. Niyogi, and M. A. Nowak. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–59, 2001.

E. H. Lenneberg. *Biological Foundations of Language*. John Wiley & Sons, Inc., New York, 1967.

M. Nakamura, T. Hashimoto, and S. Tojo. The language dynamics equations of population-based transition – a scenario for creolization. In H. R. Arabnia, editor, *Proc. of the IC-AI'03*, pages 689–695. CSREA Press, 2003.

P. Niyogi. *The Informational Complexity of Learning*. Kluwer, Boston, 1998.

M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.

Open Problems in the Emergence and Evolution of Linguistic Communication: A Road-Map for Research

Chrystopher L. Nehaniv
*Adaptive Systems Research Group
University of Hertfordshire
Hatfield Herts AL10 9AB
United Kingdom
C.L.Nehaniv@herts.ac.uk

Abstract

This paper surveys issues around several grand challenge problems for the understanding of the emergence and evolution linguistic communication, and discusses possible approaches. The identified problems the emergence of (1) advanced use of deixis, gesture, and reference; (2) predication; (3) negation; (4) syntactic categories; and (5) compositionality.

1 Introduction

In the last decade or so, there has been an explosion of interest in the modelling and understanding of language origins. The employment of simulation and robotic agent-based, connectionist neural network, and evolutionary techniques has provided new methods for formulating hypotheses, validating mechanisms, and selecting between alternative theories on the emergence of linguistic and language-like phenomena in controlled experimental settings that meet the scientific criteria of reproducibility. Recent work on the emergence and evolution of human language and more simple communication systems has been increasingly interdisciplinary, involving collaborations between linguists, philosophers, biologists, cognitive scientists, roboticists, mathematical and computational modellers – see e.g. research papers (MacLennan, 1992; Steels, 1995; Hashimoto and Ikegami, 1995; Arita and Koyama, 1998; Billard and Dautenhahn, 1999; Kirby, 1999; Nehaniv, 2000; Cangelosi, 2001; Steels, 2003) and interdisciplinary collections (Wray, 2002; Cangelosi and Parisi, 2002; Christiansen and Kirby, 2003).

This paper surveys some currently open problems in the emergence and evolution of linguistic communication that present grand challenges to those working in constructive aspects of the emergence of communication. In this paper, we address the programme of demonstrating mechanisms that achieve various language-like properties in computational agent and robotic models. This is not intended to be an exhaustive survey. Many important research articles and researchers could not be mentoned here. The discussion is instead indicative of current research activity (and inactivity) as regards a set of fundamental problems in the area.

We will discuss the following completely or largely open areas:

(1) deixis, gesture, and reference;
(2) predication;
(3) negation;
(4) emergence of syntactic categories
(5) compositionality

The emergence and modelling of these phenomena are discussed in the context of embodied, social interaction and evolution (cultural or otherwise). Ideally, mechanisms based on sensorimotor and experiential grounding in bottom-up, agent-centered models involving populations of agents will help yield deep understanding of the emergence of the above phenomena.

One area is conspicuously missing from the above list:

(0) grounding and shared vocabularies

and will also be discussed briefly below. This area has not been included in the list of current grand challenges since there has been substantial progress in it. However grounding and shared vocabularies will need to be integrated with the answers to the grand challenge problem areas (1-5) to yield grounded and shared language-like communication systems with much more complex types of vocabulary with grounded meaning than what has been so far achieved.

2 What is Meaning and What is Language For?

We regard linguistic and language-like communication as the capacity of an agent to influence the world around by the systematic use of signals mediated by their reception by other agents in its environment. Thus, language is regarded as a means for the agent to 'manipulate' the world around for its own benefit, similar to other traits of biological organisms (cf. the discussion of the transition to language from a biological viewpoint in (Maynard Smith and Szathmáry, 1995)). As Wittgenstein (1968) taught us, the *meaning* of any signalling behaviour, such as in language, arises in how it is used by the agent to manipulate its environment (including other agents) in its interactions with other agents. This can be related to the utility to an agent (in a statistical sense) of information in a signalling channel (see Nehaniv (1999); Nehaniv et al. (1999, 2002)). According the insights of Peirce (1839-1914) [republished in (Peirce, 1995)], the relationship between signs and significations is mediated by an interpretant, and the mapping between signs and what they signify is a *process* that depends on the particular agents involved and on their situated contexts. The ideas just presented follow the discussion of Nehaniv (1999, 2000). The Wittgensteinian-Peircian viewpoint outlined by Parisi et al. (2002) is similar.

In particular, these realizations lead a tremendous amount freedom in the emergence of language-like phenomena that has often been ignored and oversimplified by naively, often unconsciously, applying constraints on simulation models. This freedom and the related lack of constraints is illustrated by several corollaries. Understanding the emergence of meaning and language requires the generative synthesis of the phenomena in question beginning with the following facts:

1. Meaning is always agent-specific.

2. There is no privileged set of pre-existing space of possible meaning, containing ideal concepts.

3. There is no unique and no pre-existing syntactic structure on possible meanings.

4. If meanings, spaces of meaning, or syntax in meaning space do arise, they will be agent-specific as well.

5. The mappings between signs and meaning are mediated by interpreted signals between agents, and these mappings are also agent-specific and depend on the context of the interaction.

See (Nehaniv, 1999, 2000) for further discussion of these points.

Note that none of the above discussion refers to truth values or truth conditions, which are highly derived properties of human linguistic behaviour (Nehaniv, 2000), and that therefore should not be the starting point for an attempt to understand meaning, communication, and language. The highly refined formal tools mathematics and logic – including truth values, predicate logic, context-free grammars, denotational semantics, etc. – have allowed scientists achieve precision and thus escape from ambiguities and dependence on context and specific agents. But specific agents and context are inherent to the emergence of language, while these tools are based on abstractions and refinements from human language. Any explanation of the emergence of language that uses them as primitives to derive the phenomena that they are based on thus puts the proverbial cart before the horse (Nehaniv, 2000; Milikan, 2004).

This is not to say that these tools and formalisms should never be used. In computational modelling this is clearly would not be possible, simply due to the use of computers. No simulation or robotic study in the emergence and evolution of linguistic communication has been able to proceed successfully without simplifying some (or sometimes all) of the above complexity away. If agents are endowed with some of these language-like capacities, it is important to keep track of which ones. If new phenomena then emerge, one has an argument that the built-in capacities provide scaffolding for the new phenomena. For instance, the work of Kirby (1999) shows that, in populations of agents *with the capacity to use and derive context-free grammars*, processes of self-organization resulting from attempts to learn grammar based on induction from the evidence of grammar-generated utterances of other agents lead over generations to increasingly compositional grammars. His work does not how it is that context-free grammars nor the capacity for compositionality could first emerge (since these are given at the start).

3 Symbol Grounding & Shared Vocabularies

Different aspects of *symbol grounding* (Harnad, 1990) and the self-organization and maintenance of *shared vocabularies* are increasingly well-studied and coming to be understood, especially for vocabularies to identify or name objects (selecting one target of reference from an environment) or label situations (MacLennan, 1992; Steels, 1995, 1998; Billard and Dautenhahn, 1999; Baillie and Nehaniv, 2001; Parisi et al., 2002). Less work has been done on the grounding of shared vocabularies with more complexity, e.g. in which various parts of speech exist (labelling for example actions or actions on objects, or with compositional syntax), although the work of Cangelosi and collaborators has moved in this direction (e.g. Parisi et al. (2002)).

4 From Deixis, Gesture, and Manipulation to Reference

The items, deixis and gesture, in challenge area (1) are clearly related and emergence of reference. Reference is often suggested to be grounded in deixis and gesture but just how this occurs needs elucidation. Pointing, deictic gaze, joint attention, and gesture play important roles in the development of intersubjectivity and language in humans (cf. Kita (2003)).

Pointing, since it can be directed at many things and since it directs others' attention at them, could have provided for a kind "ur-pronominalization" in the emergence of linguistic communication. That is, pointing provides for a variable or variables that can be bound to object and persons in the environment, giving at least of degree shared reference via shared attention.

Rizzolatti and Arbib (1998) present a hypothesis on the emergence of language based on mirror-neurons in primates and humans. These neurons in the premotor cortex fire both when carrying out and when seeing an action performed. It is argued that this provides a substrate on which shared meaning can arise, as similar affordant gestures (e.g. manipulations such as grasping a fruit) are immediately understood by a conspecific interaction partner. Gestural language is then hypothesized to have developed and eventually to have given way to vocal language. Hurford (2004) acknowledges a possible role for mirror neurons in understanding the possible emergence of language, but surveys many gaps that remain in such an explanation, such as explaining the well-known arbitrariness of the sign in regard to its reference.

Milikan (2004) has a more general notion of reference that relates to utility of information in internal states or signalling channels. A more general notion of gesture regards gesture as the signalling of such useful information. This is similar to the viewpoints on the meaning of signals in (Nehaniv, 1999, 2000; Wittgenstein, 1968).

The issues discussed in this section evidently relate closely to the grounding of symbols and the emergence of shared systems of communication. Despite progress in these areas, constructive studies linking deixis and gesture to these problem areas remain to be carried out constructively in robotic and simulation models (but see Baillie and Nehaniv (2001); Baillie et al. (2004) for some first work in this direction).

5 Predication

For detailed analysis of predication and its complex structure in human language from the viewpoint of linguistics, see (Napoli, 1989). In human language, a rudimentary function of noun phrases is to pick out objects of reference from the environment (possibly even absent ones). Adjectives constrain the selection by imposing conditions on which object might be referred to.

One formal view of reference (implicit e.g. in (Steels, this volume) and classical box-world natural language processing systems) is that instances of lexical items such as a noun ("ball") or adjective ("red") are understood as predicating properties of object variables. Selection of referents is determined by solving constraints on such predicates over a space of objects in the environment. For example, $ball(X)$ and $red(Y)$, restricts the reference to a red ball if X must equal Y, as it must in the phrase "the red ball". Similarly verbs provide another class of predicates which might take multiple semantic role arguments expressed in a given syntactic subcategorization frame that resolves variable references (Steels, this volume).

As mentioned above, predicate logic and first-order logic are abstractions from the predicate structure of natural language. With the approach just described, predication itself is a primitive and therefore does not emerge. However, *a transition from reference to predication* is suggested by an association that tends to identify referential variables in one-place referential predicates (like $red(X)$), or by grammatical rules that force the identification of variables in the referential predicates.

Scenario for the Emergence of Predication.
Early on proto-words or gestural signs could have their referents *associated* in a general way, non-specific way merely by co-occurring close together in time. We elaborate a suggestion on the earliest source of predication: it may be a highly derived form of *topic-comment* structure, which is itself founded on association (Nehaniv, 2000). For instance, deictic gesture serves to select a target of joint attention (*topic*), and then another gesture or utterance near to it in time serves to communicate content that was associated to the topic as *comment*. Eventually ritualization of such communicative practice produces gramaticalization of a topic-comment construction. Predication then arises via grammaticalizaton of the special case in which not only an association between topic and comment occurs, but the comment gives to the topic a labelling category: "This - food", a property label "This – bad", or a semantic action-role "This – eat". Thus there is a progression in the emergence of predication from association and topic-comment via ritualization to grammaticalizaton of predication.

Ritualization is well-known in animal communication systems (Smith, 1977, 1996; Bradbury and Vehrencamp, 1998) and one instance of it is grammaticalization, a well-recognized process in human language change (e.g. (Bybee et al., 1994)). A clear path for research into this open area would be to proceed to validate this proposed scenario by building computational or robotic realization and showing whether and how the transitions

association $\rightarrow$ topic-comment $\rightarrow$ predication

could occur (ideally including *grounded referencing*). This should shed light on the details of the emergence of predication and the mechanisms required for this to occur.

If this could be done, more complex predication and modification could then be addressed. In more complex human language, both predicates and modifers occur. Predicates tend to mark more highly salient assertions, while modifiers tend to act in the background to tune reference via constraints (Nehaniv, 1987).

Let us again remark about the at best low relevance of truth values here. In early language as in animal communication system, the emphasis was of course manipulation of and influence in the environment via signalling to others (cf. Maynard Smith and Szathmáry (1995); Milikan (2004); Nehaniv et al. (2002)), rather than on propositional assertions. Truth values of predicates on objects was only a later invention and abstraction of humans.

6 Negation: A Small Research Programme

It seems little has been done in emergence of negation in constructive evolution of language models. A discussion of negation of speach acts and within speech acts occurs in (Searle, 1980/1969). A comprehensive book on negation is (Horn, 2001).

Early Scenarios for Negation.
(The material in this subsection is modified from text by Donna Jo Napoli (Napoli, pers. comm.).) Early predicates used by early humans likely indicated actions such as "come", "hide", "be quiet", "run", or referenced objects, such as "food", "water". Negation can operate on nouns as well as on verbs, or other parts of speech, and is, of course, a predicate in itself.[1] Letting others know there is nothing in the cave, for example, was probably a pretty important early message. So one would expect "nothing" or "no living thing" to be an early negation.

Non-verbal, facial and manual gestures may have played an important role in early negation. When hunting, when trying to be quiet for any reason, people have always used their faces and hands. We all recognize the hush gesture. We know to raise our eyebrows to ask yes/no. This sort of thing is extremely common around the world. In Australia, many tribes used to have sign languages just for hunting. (They had sign languages for other things, too – like to use with widows – and for the deaf). The first negation was likely either facial or gestural – perhaps a head shake or lowered brows (as in American Sign Language (Neidle et al., 2001)), or protruded lips. Also, early negation was likely simultaneous with whatever was being negated, whether spoken words or other gestures. So shake your head and say "buffalo" - or shake you head and say "swim/enter water" or shake your head and gesture "walk (whatever that gesture might be for those peoples) – and you're getting across the messages "there are no buffaloes" – "don't go in the water" – "don't walk". (Scenario and examples due to Napoli (pers. comm.). The author is responsible for any misrepresenations of her views.)

[1] Or a modifier, where modification is plays a role, e.g. a specifying a constraint on reference within consituent syntactic structure, and is generaly less marked than predication.

Computational Scenarios for Emergence of Negation.

We now give several ideas for constructivist approaches to negation:

1. It seems straight forward to use inhibition in artificial neural networks to suppress the behaviour in the presence of a negation signal N. Suppression of all action could yield compliance (by inaction) with commands such as "don't touch that". This could be realized to many existing models.

2. A research scenario into the use of more specific negation could employ connectionist neural network models of agents using linguistic signalling such as those of Cangelosi (2001), which can have a noun-verb distinction (see below) that they exhibit in language games. We propose that these be extended by the introduction of tasks into the language games that sometimes involve negation: When the new signal N co-occurs with a previously learned linguistic signal S the language game task requires choosing a *different* object/property ("(proto)noun"/"(proto)adjective") or action ("(proto)verb"), respectively, than would be for the signal S. Tasks without the signal N must also be carried out by the agents and require the original interpretation of S. That is, the agents could carry commands such as "pull cup", "not-pull [e.g. push] cup", "pull not-cup" (i.e. pull an object other than the cup), , or even "not-pull not-cup" (e.g. pushing a ball would be a correct response). The meaning of the negation signal N would be grounded in the language game tasks these agents have to perform. Demonstrating that evolving populations of neural network agents could learn this task would establish a connectionist basis for specific negation of constituents of simple linguistic utterances. Alternatively, one could do the same kind of study using agents such as in the work of Steels (2003).

3. We note that in many human sign languages such as American Sign Language (ASL), the scope of negation can be given over syntactic subunits by non-manual gestures. In ASL non-manual marking (furrowing of the eyebrows and side-to-side headshake) may spread over the (c-command) domain of constituent syntactic node, and moreover such spread is obligatory in the absence of manual marker (Neidle et al., 2001). (This property agrees well with the likely simultaneity in the early negation described in Napoli's scenario above.) Thus in constructivist studies of the emergence of language, it would be very interesting to investigate scope of negation. For example, in neural network agent models, the use of a negative signal would have presumably to involve the persistence in the network of internal state over the scope of the negated constituent. Synthetic neural imaging techniques like those of Cangelosi and Parisi (2004) could be useful here.

7 Syntactic Categories

In artificial neural network connectionist models, Parisi et al. (2002) have shown the grounded emergence of rudimentary nouns and verbs: Nouns, as linguistic signals that co-vary with sensory stimuli, and verbs, as linguistic signals that co-vary with actions (largely independent of sensory stimuli). They have suggested that this could be extended to (proto)adjectives, that select a referent within a noun category using some intrinsic property, and to non-adjectival modifiers, such as location indicators (e.g. left, right, above), that reflect more temporary properties of objects which are not instrinsic to the object but depend on the relationship of object to speakers and the environment. This remains to be done, as does increasing the complexity of syntactic categories the approach can generate (e.g. to verbs with a patient and recipient role, as "give the apple to Mary".

Steels (this volume) also considers the emergence of shared semantic and syntactic frames based on grammaticalization driven by computational needs of disambiguation.

This issue of emergence of syntactic categories, which are restricted in the types of semantic environments where they can occur (as in the work of Parisi et al. (2002)), and in their signal contexts, and in the types of arguments they can take (if any), leads to the next grand challenge, the achievement of full-blown compositional syntax in a grounded communication system.

8 Compositionality

The emergence of lexical items that take arguments (such as transitive verbs that take an noun-phrase as object) is called compositionality. This has syntactic and semantic aspects, and accounts for much of the combinatorial richness of human language. There have also been a growing number of studies on the emergence of various aspects of syntax (e.g. (Kirby, 1999; Cangelosi, 2001; Steels, this volume)). While there has also been some pioneering work on syntactic categories (e.g. Cangelosi and Parisi (2004); Parisi et al. (2002)), and grounded compositionality (Steels, 1998), many aspects of compositionality in linguistic communication remain completely open for constructive modellers to begin to explain.

Segmentation and pauses in modern human speech, e.g. arising from the need to breath or the temporal nature of cognitive processes, combined with local context have been shown information-theoretically to improve the disambiguation of speech, suggesting that sequential process of smaller sequential units may help provide the basis for syntax in language evolution and language processing Lyon et al. (2003).

Cangelosi (2001) showed the emergence of verbs for actions that take target objects references in neural network agents that can manipulate simple objects in the environment in an evolutionary simulation, but non-compositional communicative signals could also evolve.

Assuming a fixed and syntactically structured meaning space, and a capacity to use and learn context-free grammars, Kirby (1999, 2001), as mentioned above, has shown that grammars with high degrees of compositionality arise and are easier to transmit over the course of generations of learning in such agents starting from agents using non-compositional 'holistic' grammars (i.e. with a different utterance for each meaning). Extending this work to agent-centred spaces of meaning grounded in interaction and language games remains to be achieved.

Steels (this volume) argues that the purpose of compositional grammar is to reduce the number of variables in a decoded meaning structure in order to hope with compuational complexity in interpretation. He constructs agents in simulation studies that apply this principle and are able to converge on shared grammars by reinforcing and modifying syntatic and semantic role-structural frames (to propogate referental constraints) based on communicative success and failure. The same structures are used for parsing and for production.

Recursive composition structure is possible if the expansion of argument can non-trivially include the same argument type (as with clauses embedded in other clauses). When this occurs, in principal the language becomes unbounded in size.

9 Conclusions

Our list of grand challenge areas identified five challenges beyond symbol grounding and the emergence of shared systems of communicating meaning: (1) the role of deixis, gesture, and manipulation in the grounding and emergence of reference, (2) predication, (3) negation, (4) syntactic categories, and (5) full syntax - compositionality and recursive structure.

Challenges (2), on predication, and (3), on negation, have been the most neglected by the evolution of language community. We hope this paper stimulates discussion on these issues and promote research especially into those areas.

The problem of predication (2) is argued to be related to associative processes and to topic-comment structures, as precursors. Predicates as the exist today in human languages are seen as a highly derived special case of related processes.

Computational scenarios for studying the emergence of predication and of negation have been proposed and discussed in order to encourage the investigation into their hemergence.

Other immediate work to be done to meet these grand challenges includes: (4) emergence of syntactic categories needs to be shown without assuming an underlying categorization on some pre-existing space of meanings in grounded language games. (5) compositionality (and recursion) needs to be shown to emerge in a setting of grounded meaning without the assumption of an underlying grammatical ability, such as the capacity to learn and use context-free grammars.

Acknowledgements

Thanks to Donna Jo Napoli for many ideas relevant to the early negation and for discussing some thoughts on early predication.

References

Takaya Arita and Yuhiji Koyama. Evolution of lingusitic diversity in a simple communication system. *Artificial Life*, 4(1):109–124, 1998.

Jean Baillie and Chrystopher L. Nehaniv. Deixis and the development of naming in asynchronously interacting connectionist agents. In *First International Workshop on Epigenetic Robotics*, pages 123–129. Lund University Cognitive Studies, vol. 85, 2001.

Jean Baillie, Chrystopher L. Nehaniv, Patrick Quick, Attila Egri-Nagy, and Sandra Warren. Deixis, interaction topology, and the emergence of naming. In *First International Workshop on the Emergence and Evolution of Linguistic Communication*, pages 33–40. Japanese Society for Artificial Intelligence, 2004.

A. Billard and K. Dautenhahn. Experiments in learning by imitation - grounding and use of communication in robotic agents. *Adaptive Behavior*, 7 (3/4), 1999.

Jack W. Bradbury and Sandra L. Vehrencamp. *Principles of Animal Communication*. Sinauer, 1998.

Joan Bybee, Revere Perkins, and William Pagliuca. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. University of Chicago Press, 1994.

Angelo Cangelosi. Evolution of communication using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*, 5(2):93–101, 2001.

Angelo Cangelosi and Domenico Parisi, editors. *Simulating the Evolution of Language*. Springer Verlag, 2002.

Angelo Cangelosi and Domenico Parisi. The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging. *Brain and Language*, 89(2):401–408, 2004.

Morten H. Christiansen and Simon Kirby, editors. *Language and Evolution*. Oxford, 2003.

Stevan Harnad. The symbol grounding problem. *Physica D*, 42, 1990.

Takashi Hashimoto and Takashi Ikegami. Evolution of symbolic grammar system. In F. Morán, A. Moreno, J. J. Merolo, and P. Chacón, editors, *Advances in Artificial Life (3rd European Conference on Artificial Life, Granada, Spain, June 1995)*, volume Lecture Notes in Artificial Intelligence, 929, 1995.

Laurence R. Horn. *A Natural History of Negation*. CSLI Pubilications (The David Hume Series: Philosophy and Cognitive Science Reissues), 2001.

James Hurford. Language beyond our grasp: What mirror neurons can, and cannot, do for the evolution of language. In Kimbrough Oller and Ulrike Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 297–313. MIT Press, Cambridge, MA, 2004. URL `http://www.ling.ed.ac.uk/~jim/mirror kl.s.ps`.

Simon Kirby. Learning, bottlenecks, and infinity: a working model of the evolution of syntactic communication. In K. Dautenhahn and C. L. Nehaniv, editors, *Proceedings of the AISB'99 Symposium on Imitation in Animals and Artifacts*, pages 55–63. Society of the Study of Artificial Intelligence and the Simulation of Behaviour, 1999.

Simon Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001.

Sotaro Kita, editor. *Pointing: Where Language, Culture and Cognition Meet*. Lawrence Erlbaum Associates, Inc, 2003.

Caroline Lyon, Chrystopher L. Nehaniv, and Bob Dickerson. The segmentation of speech and its implications for the emergence of language structure. *Evolution of Communication*, 4(2):161–182, 2003.

Bruce MacLennan. Synthetic ethology: An approach to the study of communication. In *Artificial Life II*. Addison Wesley, 1992.

John Maynard Smith and Eörs Szathmáry. *The Major Transitions in Evolution*. W.H. Freeman, 1995.

Ruth Garrett Milikan. *Varieties of Meaning - The 2002 Jean Nicod Lectures*. MIT Press, 2004.

Donna Jo Napoli. *Predication theory: A Case Study for Indexing Theory*, volume 50. Cambridge Univ. Press, 1989.

Donna Jo Napoli. email correspondence, February 2005, pers. comm.

Chrystopher L. Nehaniv. Predication and modification in Chinese and universal grammar. Honors Thesis in Linguistics, University of Michigan, 1987.

Chrystopher L. Nehaniv. Meaning for observers and agents. In *IEEE International Symposium on Intelligent Control/Intelligent Systems and Semiotics (ISIC/ISAS'99)*, pages 435–440, 1999.

Chrystopher L. Nehaniv. The making of meaning in societies: Semiotic & information-theoretic background to the evolution of communication. In *Starting from Society*, pages 73–84. Society for the Study of Artificial Intelligence and Adaptive Behaviour, 2000.

Chrystopher L. Nehaniv, Kerstin Dautenhahn, and Martin J. Loomes. Constructive biology and approaches to temporal grounding in post-reactive robotics. In *Sensor Fusion and Decentralized*

Control in Robotics Systems II (September 19-20, 1999, Boston, Massachusetts), Proceedings of SPIE Vol. 3839, pages 156–167, 1999.

Chrystopher L. Nehaniv, Daniel Polani, Kerstin Dautenhahn, René te Boekhorst, and Lola Ca namero. Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In *Artificial Life VIII*, pages 345–349. MIT Press, 2002.

Carol Neidle, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. MIT Press, 2001.

Domenico Parisi, Angelo Cangelosi, and Ilaria Falcetta. Verbs, nouns and simulated language games. *Italian Journal of Linguistics*, 14(1):99–114, 2002.

Charles S. Peirce. *Collected Papers, Volume 2: Elements of Logic*. Harvard, 1995.

G. Rizzolatti and M. A. Arbib. Language within our grasp. *Trends in Neurosciences*, 21(5):188–194, 1998.

John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1980/1969).

W. John Smith. *The Behavior of Communicating: An Ethological Approach*. Harvard, 1977.

W. John Smith. Communication and expectations: A social process and the operations it depends upon and influences. In Marc Bekoff and Dale Jamieson, editors, *Readings in Animal Cognition*, pages 243–255. MIT Press, 1996.

Luc Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2:315–332, 1995.

Luc Steels. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103:1–24, 1998.

Luc Steels. Evolving grounded communication for robots. *Trends in Cognitive Science*, 7(7):308–312, 2003.

Luc Steels. What triggers the emergence of grammar? In Angelo Cangelosi and Chrystopher L. Nehaniv, editors, *Seconnd International Symposium on the Emergence and Evolution of Linguistic Communication*. Society for the Study of Artificial Intelligence and Adaptive Behaviour, this volume.

Ludwig Wittgenstein. *Philosophical Investigations (Philosophische Untersuchungen)* – German with English translation by G.E.M. Anscombe. Basil Blackwell, 3rd edition, 1968.

Alison Wray, editor. *The Transition to Language*. Oxford, 2002.

Neural substrates for segmentation and semanticization of song strings

Kazuo Okanoya[*]
[*]Labolatory for Biolinguisitcs
BSI, RIKEN, Japan
okanoya@brain.riken.jp

Bjorn Merker[†]
[†]Department of Psychology
Uppsala University, Uppsala, Sweden
gyr694c@tninet.se

Abstract

Language is a unique human behaviour. Language has the characteristics that divide itself from other animal communication systems: the combinational creativity based on limited numbers of arbitrary (socially established) tokens. Nevertheless, rudiments of some of the sub-faculties that enabled language can be seen in animals. We identified three of these non-trivial sub-faculties and considered neural substrates that enabled each of these. In Merker & Okanoya (in this volume) we presented a scenario for string semanticization by merging of songstrings and situational meanings. Based on the neural substrates assumed here, we propose a possible neural mechanism of language emergence.

1 Introduction

Birds and whales, as well as humans, can learn strings of vocal utterances. They even put tokens into a syntactical form (Okanoya, 2002, 2004; Payne, 2000). Parrots, chimps, and monkeys, as well as other animals, can learn to associate situational meaning and particular sound (Pepperberg, 1999; Cheney & Seyfarth, 1990). But humans only can use language, a set of semanticized strings that are acquired through leaning (Fig. 1).

Thus, the Rubicon that devides us and other animas is the semanticization of vocal strings. In an associated abstract, Merker & Okanoya (2005) provided evolutionary scenario for string semanticization and an origin of language. Here we suggest possible neural substrates that enable the process of string semanticization (Fig. 2).

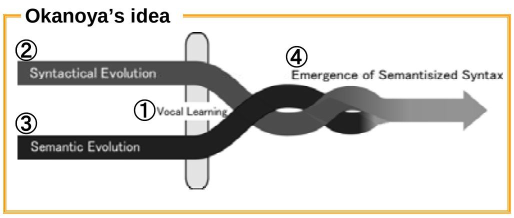


Figure 1: Independnt evolution hypothesis put forward by Oknanoya (2002). 1. Vocal learning is the basis of linguistic faculty. 2. Syntactical evolution occurred as a courtship ritual that preferred for behavioral complexity. 3. Semantic evolution occurred as a kin-selection and reciprocal altruism in a lasting society. 4. Emergence of semanticsized syntax occurred by mutual segmentation of strings and context.

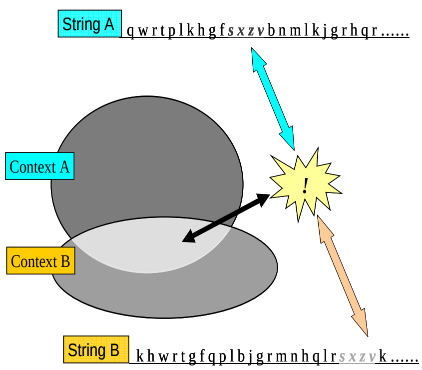


Figure 2: Mutual segmentation of songstring and behavioural context. Songstring A is sung in context A and songstring B in context B. The common part of the context A and B is associated with the common sub-string of the string A and B. Based on the idea by Bjorn Merker.

2 Neural substrates

2.1 Vocal learning

Vocal learning independently evolved several times in vertebrates (Payne, 2000; Catchpole & Slater, 1995; Boughman, 1998). Vocal learning refers to motor learning of spectro-temporal features of sound signals that are used in intra- and inter species communication by conspecific members. Most animal vocalizations are innately determined. Vocal learning, experience dependent, categorical, and long-lasting modification of vocal output, do not occur in most animals and only a few species of animals including whales, bats, birds, and humans exhibit vocal learning.

Are there any specific anatomical substrates that correlate with the faculty of vocal learning? One of the candidates for this question is the direct cortical-medullar pathway for articulation and breathing. In humans, a part of motor cortex directly projects to the medullary nuclei, the nucleus ambiguus and the nucleus retro-ambiguus (Kuypers, 1958). This projection is absent in the squirrel monkey and chimpanzee. Jurgens (2002) thus assumes that this projection exists only in humans among primates. Similarly, there is a direct cortical-medullar pathway for articulation and breathing in the zebra finch, a species of songbirds, but a similar projection in pigeons do not exist. Most of pigeon vocalizations are considered to be innate (Wild, 1993; Wild et al., 1997). Considering these evidence, we can hypothesize that this projection exists in the species that show vocal learning while it is absent in the species without vocal learning (Fig. 3).

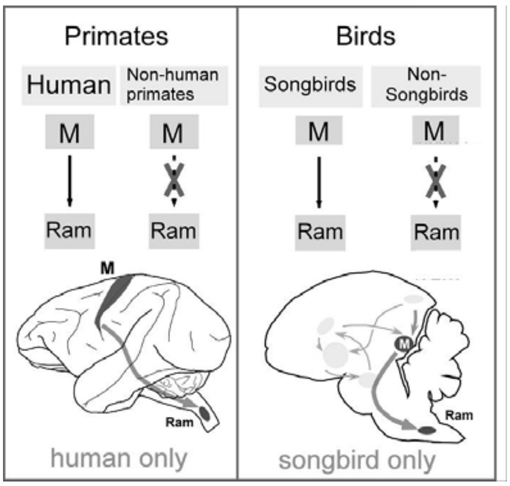


Figure 3: Direct motor cortex – medullary nuclei related with breathing exist only in humans, songbirds and whales (Okanoya et al., 2004).

While this projection exist only in a limited number of species, it may be possible, that this projection is simply very faint in most of species. Deacon (1997) introduced anecdotal story of a zoo seal that learned to mimic speeches of drunken persons (Deacon, 1997). This seal had a brain inflammation as young and Deacon suspected that during the process of recovery the cortical medullar projection might be reinforced in this particular animal. If such cases could occur, by training animals to perform spontaneous vocalization while they are young, we may be able to reinforce this pathway and induce vocal learning in a species that was said to be non vocal learners.

2.2 Syntax

Songbirds learn courtship songs from adult males (reviewd in Catchpole & Slater, 1995; Zeigler & Marler, 2004). They often splice parts of songs that are sung by different males (Fig. 4). Birds probably do so by using statistical information in conspecific songs: chunking of song notes occurs at junctions of lower transition probability.

In humans, a phrase structure is processed as a perceptual unit. Perceptual position of an embedded click moves outside of the phrase structure (Foder & Bever, 1965). We examined whether similar phenomenon could be observed in songbirds.

We trained Bengalese finches in a click-detection experiment. When a background of his own song was played in the test box, reaction time of the subject bird to detect the click was longer than without the background song or with the background song played in reverse: detection of clicks is postponed until a chunk of song notes is processed. Thus, chunks of birdsong, like chunks of linguistic elements, are processed as a cognitive unit (Suge & Okanoya, submitted).

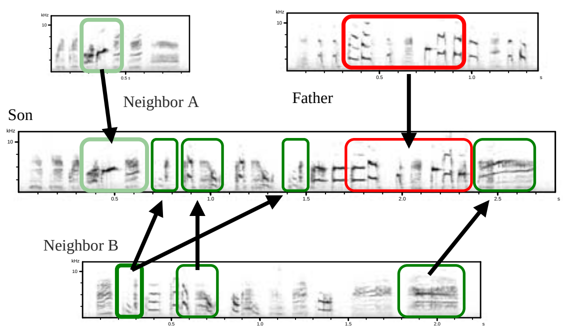


Figure 4: An example of perceptual segmentation

observed in a Bengalese finch son reared in a multi-family environment (Takahashi & Okanoya, unpublished data).

The auditory segmentation observed in song perception maybe governed by cortex-basal ganglia loop. of auditory. When a part of basal ganglia was destroyed in adult male Bengalese finches, patterns of song note segmentation changed (Fig. 5).

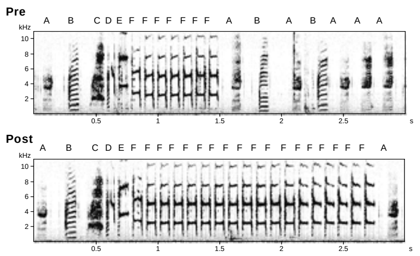


Figure 5: Deterioration of motor chunking of song notes in Bengalese finch after the basal ganglia lesion (From Kobayashi et al., 2001).

In an ERP study with human subjects, we found that string segmentation was associated with a negative brain potential that was strongest at electrode placed nearby the anterior cingulate cortex, suggesting that string segmentation involves the cortex-basal ganglia loop. The same loop is known to be used birdsong learning. Thus, segmentation of birdsong strings may also be governed by the cortex-basl ganglia pathway (Fig. 6).

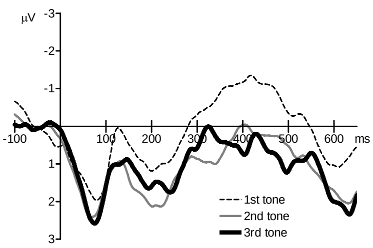


Figure 6: Average ERP recorded from the electrode situated at the point reflecting the activity of the anterior singulate cortex. Three-tone words were randomly defined and these words were randomly stringed and presented to the subject. In a three-tone word, the second and third tone could be predicted by what comes for the first, but the first tone could not because words were randomly stringed. Dotted line indicated the ERP recorded after the onset of 1st tone, showing a specific process presumably reflecting the errors in sequence estimation (Abla et al, unpublished data).

2.3 Semantics

Behavioural contexts could be associated with particular behaviour tokens thus forming a rudiment of "naming" in several animals. We observed that when trained to use a rake to retrieve a distant food, monkeys be-gan to vocalize "coo" calls spontaneously. They did so especially when the prepara-tion of the rake tool by the experimenter delayed (Hihara et al., 2003). To further investigate this phenomenon, we systematically manipulated behavioral contexts by presenting the tool or food whenever the monkey made a vocalization regardless of the types of the calls. In one experimental situation, the experimenter gave a food at a distance when the monkey produced a coo call (Call A). By the second coo call (Call B) the experimenter placed a rake tool to the monkey. The monkey could retrieve the food by the rake. In another, the experimenter gave the rake to the monkey beforehand. A food was placed at a distance when the monkey vocalized a coo call (Call C). Likewise, we never tried to differentiate the calls After 5 sessions of trainings, the monkeys eventually used acoustically distinct types of calls when they asked for the tool (Call 2) or food (Call 1 and 3). The calls used to ask for the tool was longer and higher pitched than the ones used to ask for the food.

We argue that the different reward conditions (food or tool) set up different emotional contexts for the monkeys. Different emotional contexts, in turn, affected the production of coo calls differently for the tool or food situations. Since the tool train-ing activated the neo-cortex very highly, the calls were associated with different behavioural contexts. Thus, the calls became categorized and emotionally differenti-ated calls gradually became categorical vocalizations. Through this process, we suspect the emotional coo calls changed into categorical labels denoting the behavioural situation.

We speculate this categorization of vocal tokens maybe related with highly specified behavioural situations. Such specified behavioural situations would evoke specific emotional content in limbic system including amygdala and hippocampus. The sates of the excitation in the limbic system maybe labeld by the emotional vocalizations associated with the situations.

3 Merging of syntax and semantics

Taking these speculations together, we propose that longer songstrings are segmented by the action of the cortex-basal ganglia pathway and each part-string is represented by a menmonc.

A mnemonic can then be associated with particular behavioural contexts segmented by the pre-frontal cortex-hippocampus- amygdala loop. The parallel operation of these two systems would then enable mutual segmentation and matching of behavioural context and song strings, a process tantamount to an incipient language system (Merker & Okanoya, 2005).

Although no experimental evidence is available at this point, experiment directly asking the above hypothesis could be designed and performed. In an exploration of neural correlates of list learning, Christie &Dalrymple-Alford (2004) found that only learning of a short (4-item) list was retained after basal ganglia lesions but learning of longer lists (8 or 12 items) were affected by such lesions. On the other hand, lesioning the hippocampus did not produce any notable effects on the list learning. Although results are contrary to what we expect from the above hypothsis, similar paradigm with longer list length and more elaborated training might produce different results.

Thus language maybe possible without assuming a special "recursion" device as suggested by Hauser, Chomsky and Fitch (2002). Recursive function might arise secondarily from the interaction between the cortex-basla ganglia loop and the prefrontal/ hippocampal/ amygdale loop (Fig. 7).

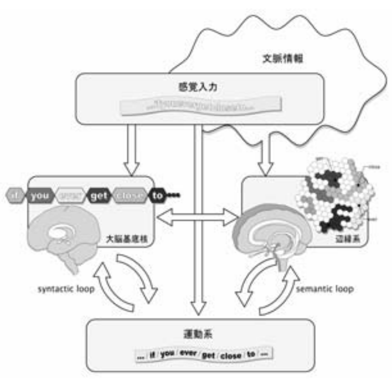


Figure 7: Neural substrates for mutual segmentation of behavioral context and songstring. A long string can be segmented by the statistical learning of the cortex-basal ganglia system. Simultaneously, behavioural contexts can be segmented by the pre-frontal cortex /amygdala/hippocampus system. A segmented sub-string maybe labelled by a behavioural context and given a particular behavioural token to it.

Acknowledgements

We thank Maki Ikebuchi, Miki Takahashi, and Yasuko Tobari for inspirations. The preparation of the manuscript was supported by Japan Science and Technology Agency.

References

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proc. R. Soc. Lond. B*, 265, 227-233.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge, UK: Cambridge University Press.

Cheney, D. L. & Seyfarth, R. M. (1990). How monkeys see the world. University of Chicago Press, Chicago.

Christie &Dalrymple-Alford (2004) A new rat model of the human serial reaction time task: contrasting effects of caudate and hippocampal lesions. *J Neurosci.,* 24, 1034-9.

Darwin C. (1871). *The descent of man and selection in relation to sex.* D. Appleton & Company, New York.

Deacon, T. W. (1997). *The symbolic species.* Norton & Company, New York.

Fodor, J, & Bever, T (1965). The psychological reality of linguistic segments. *J verbal learn verbal behav* 4, 414-420.

Hauser, M., Chomsky, N. & Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.

Hihara S, Yamada H, Iriki A, Okanoya K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neurosci Res.* 45, 383-9.

Jurgens U. (2002). A study of the central control of vocalization using the squirrel monkey. Med. Eng. Phys., 7-8 , 473-477.

Kobayashi, K., Uno, H. & Okanoya, K. (2001). Partial lesions in the anterior forebrain pathway affect song production in adult Bengalese finches. *NeuroReport*, 12, 353-358.

Kuypers, H. G. J. M. (1958). Corticobulbr connection to the pons and lower brain-stem in man. Brain, 81, 364-88.

Merker, B. & Okanoya, K. (2005). Contextual semanticization of song strings syntax: a possible path to human language. EELC05 abstract.

Okanoya, K. (2002). Sexual Display as a Syntactic Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection. In A. Wray (Ed.), *The transition to language*, pp. 44-64. Oxford: Oxford University Press.

Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. Advances in the study of behaviour, 34, 297-345.

Okanoya, K., Hihara, S., Tokimoto, N., Tobari, Y. & Iriki, A. (2004). Complex vocal behaviour and cortical-medullar projection. EELC04 abstract.

Pepperberg, I. (1999). The Alex studies: cognitive and communicative abilities of grey parrots. Harvard University Press, Cambridge.

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In N.L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music*, pp. 135-150. Cambridge, MA: The MIT Press.

Suge, R. & Okanoya, K. (under revision). Perceptual chunking in self-produced song by Bengalese finches. *Animal Cognition.*

Wild, J. M. (1993). Descending projections of the songbird nucleus rubustus archistriatalis. *J. Comp. Neurol.* 338, 225-241.

Wild, J. M., Li, D., Eagleton, C.(1997). Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory – vocal nuclei in the brainstem of pigeon and zebra finch. *J. Comp. Neurol.* 377, 392-413

Zeigler, H. P. & Marler, P. (Eds.) (2004). Behavioral Neurobiology of Birdsong. *Annals of the New York Academy of Sciences, 1016.*

The Self-Organized Origins of Phonotactics and Phonological Patterns

Pierre-Yves Oudeyer*
*Sony CSL Paris, 6, rue Amyot, 75005 Paris, France
py@csl.sony.fr

Abstract

In previous papers, we presented a system which showed how a society of agents could self-organize a shared discrete vocalization system, starting from holistic inarticulate vocalizations. The originality of the system was that: 1) it did not include any pressure for communication; 2) it did not include any social capacity (agents did not play a language game for example); 3) it pre-supposed neither linguistic capacities nor the existence of conventions. We present here an extension of the system which shows how rules of sound combination as well as patterns of combinations can self-organize and be shared by the society of agents. This illustrates how phonotactics might have bootstrapped.

1 Introduction

Human vocalizations have a complex organization. They are discrete and combinatorial: they are built through the combination of units, and these units are systematically re-used from one vocalization to the other. These units appear at multiple levels (e.g. the gestures, the coordination of gestures, the phonemes, the morphemes). While the articulatory space that defines the physically possible gestures is continuous, each language only uses a discrete set of gestures. While there is a wide diversity of the repertoires of these units in the world languages, there are also very strong regularities (for example, the high frequency of the 5 vowel system /e,i,o,a,u/). The way the units are combined is also very particular: 1) not all sequences of phonemes are allowed in a given language (this is its phonotactics), 2) the set of allowed phoneme combinations is organized into patterns. This organization into patterns means that for example, one can summarize the allowed phonemes of Japanese by the patterns "CV/CVC/VC", where "CV" for example defines syllables composed of two slots, and in the first slot only the phonemes belonging to a group that we call "consonant" are allowed, while in the second slot, only the phonemes belonging to the group that we call "vowels" are allowed.

It is then obvious to ask where this organization comes from. There are two complementary kinds of answers that must be given (Oudeyer, 2003). The first kind is a functional answer stating which is the function of systems of speech sounds, and then showing that systems having the organization that we described are efficient for achieving this function. This has for example been proposed by (Lindblom, 1992) who showed that discretenesss and statistical regularities can be predicted by searching for the most efficient vocalization systems. This kind of answer is necessary, but not sufficient: it does not say how evolution (genetic or cultural) might have found this optimal structure. In particular, naive Darwinian search with random mutations (i.e. plain natural selection) might not be sufficient to explain the formation of this kind of complex structures: the search space is just too large (Ball, 2001). This is why there needs a second kind of answer stating how evolution might have found these structures, and in particular, how self-organization might have constrained the search space and helped natural selection. This can be done by showing that a much simpler system spontaneously self-organizes into the more complex structure that we want to explain.

(Oudeyer, in press) has shown how a system of this kind, based on the coupling of generic neural devices which were innately randomly wired and implanted in the head of artificial agents, could self-organize so that the agents develop a shared vocalization system with discreteness, combinatoriality and statistical regularities. The originality of the system was that: 1) it did not include any pressure for communication; 2) it did not include any social capacity (agents did not play a language game for example); 3) it pre-supposed neither linguistic capacities nor the existence of conventions. We present now an extension of this system which gives an account of the formation of rules of sound combination as well as of patterns of sound combinations. This amounts to the formation of phonotactics. The extension is based on the addition of a map of neurons with temporal receptive fields. These are initially randomly pre-wired, and control the sequential programming of vocalizations. They evolve with local adaptive synaptic dynamics.

2 The system

We are going to make a summary of the architecture presented in details in (Oudeyer, in press), before presenting the extension. The system is composed of agents which are themselves composed of an artificial brain connected to an artificial vocal tract and an artificial ear. Agents can produce and hear vocalizations. As described in (Oudeyer, in press), one can model each component from the most abstract to the most realistic manner. In this paper, our goal is to explore the principles of the formation of phonotactics and of phonological patterns, rather than to build a realistic predictive model. Thus, we will use the most abstract version of the components presented in (Oudeyer, in press). In particular, this means that agents produce two-dimensional vocalizations (one articulatory dimension and one temporal dimension). We use only one space to represent vocalizations: the perceptual space is bypassed and only the motor space is used. So, we pre-suppose that agents can translate a vocalization from the perceptual space to the motor space, which is acceptable since in (Oudeyer, in press) we showed how this mapping could be learnt by the agents. The articulatory dimension that we use is also abstract, but one could imagine that it represents the place or the manner of constriction for example. Finally, the agents are put in a virtual space in which they wander randomly, and at random times they generate vocalizations which are heard by themselves as well as the closest agent.

The brain of the agent is organized into two neural maps: 1) one "spatial" neural map coding for static articulatory configurations; 2) one "temporal" neural map coding for the sequences of activations of the neurons in the static neural map (this constitutes the extension of the system presented in (Oudeyer, in press)).

2.1 The spatial neural map

The spatial neural map contains neural units N_i which have broadly tuned gaussian receptive fields. We denote $v_{i,t}$ the centre of the gaussian related to N_i, which we call its "preferred vector" since it corresponds to the stimulus which activates maximally the neural unit. If we note $G_{i,t}$ the tuning function of N_i at time t, s one input vector, $v_{i,t}$ the preferred vector of N_i at time t, then:

$$G_{i,t}(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(v_{i,t}.s)^2/\sigma^2}$$

The parameter σ determines the width of the gaussian, and so if it is large the neurons are broadly tuned (a value of 0.05, as used below, means that a neuron responds substantially to 10 percent of the input space).

All the neural units have initially a random preferred vector, following a uniform distribution. Each neural unit codes for an articulatory configuration, defined by the value of its preferred vector. If the neural unit is activated by the agent and a GO signal is sent to the neural map, then there is a low-level control system which drives the articulators continuously from the current configuration to the configuration coded by the activated neuron. A vocalization is thus here a continuous trajectory in the articulatory space, produced by the successive activation of some neural units in the spatial neural map, combined with a GO signal. As we will see later on, this activation is controlled internally by the temporal neurons.

As we explained earlier, we use only one space to represent vocalizations. Thus, when an agent produces a vocalization, defined by its trajectory in the articulatory space, the agent that can perceive this vocalization gets directly the trajectory in the articulatory space. The perception of one vocalization produces changes in the spatial neural map. The continuous trajectory is segmented in small samples corresponding to the cochlea time resolution, and each sample serves as an input stimulus to the spatial neural map. The receptive fields of neural units adapt to these inputs by changing their preferred vector (the width of the gaussian does not evolve). For each input, the activation of each N_i is computed, and their receptive field updated so that if the same stimulus comes again next time, it will respond a little bit more (this is weighted by their current activation). Basically, adaptation is an increase in sensitivity to stimuli in the environment. The formula is:

$$G_{i,t+1}(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(v_{i,t+1}.s)^2/\sigma^2}$$

where $G_{i,t+1}$ is the tuning function of N_i at time $t+1$ after the update due to the perception of s_t at time t, and $v_{i,t+1}$ the updated preferred vector of N_i:

$$v_{i,t+1} = v_{i,t} + 0.001.G_{i,t}(s_t).(s_t - v_{i,t})$$

From a geometrical point of view, the preferred vector of each neural unit is shifted towards the input vector, and the shift is higher for unit which respond a lot than for unit which do not respond very much. [1]

2.2 The temporal neural map

In (Oudeyer, in press), the production of vocalizations was realized by activating randomly neurons in the spatial map. There was no possibility to encode the order in which the neurons were activated, and as a consequence agents ended by producing vocalizations in which all phoneme combinations were allowed (but of course only the phonemes that appeared as a result of the self-organization of the neural map were used). On the contrary, we will use here a temporal neural map which can encode the order of activations of spatial neurons, and is also used to activate the spatial neurons.

Each temporal neuron is connected to several spatial neurons. A temporal neuron can be activated by the spatial neurons through these connections. The tuning function of temporal neurons has a temporal dimension: their activation depends not only on the amplitude of the activation of the spatial neurons to which they are connected, but depends also on the order in which they are activated, which itself depends on the particular vocalization which is being perceived. The mathematical formula to compute the activation of the temporal neuron i is:

$$GT_i =$$

$$\sum_{t=0}^{T}\sum_{j=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}.e^{\|t-T_j\||^2/\sigma^2}.\frac{1}{\sqrt{2\pi}\sigma}.e^{\|G_{j,t}\||^2/\sigma^2}$$

with T denoting the duration of the perceived vocalization, N the number of spatial neurons to which it is connected, T_j a parameter which determines when the temporal neuron i is sensitive to the activation of the spatial neuron j, and $G_{j,t}$ the activation of the spatial neuron j at t. Here, the T_j values are such that the temporal neuron that they characterize is maximally activated for a sequence of spatial neuron activation in which two neurons are never maximally activated at the same time and for which the maximal activation is always separated by a fixed time interval. In brief, this means that rhythm is not taken into account in this simulation: we just consider order. Mathematically,

$$T_0 = 0, T_1 = \tau, t_2 = 2.\tau, ..., T_N = N.\tau$$

where τ is a time constant.

As stated in the first paragraph, the temporal neurons are also used to activate the spatial neurons. The internal activation of one temporal neuron, coupled with a GO signal, provokes the successive activation of the spatial neurons to which it is connected, in the order specified by the T_j parameters. This implies that the temporal pattern is regular, and only one neuron is activated at the same time. In this paper, each temporal neuron will be connected to only two spatial neurons, which means that a temporal neuron will code for a sequence of two articulatory targets ($N = 2$). This will allow us to represent easily the temporal neural map, but this is not crucial for the results. When an agent decides to produce a vocalization, which it does at random times, it activates one temporal neuron chosen randomly and sends a GO signal.

Initially, a high number of temporal neurons are created (500), and are connected randomly to the spatial map with random values of their internal parameters. Using many neurons makes that basically all possible sequences of activations of spatial neurons are encoded in the initial temporal neural map. The plasticity of the temporal neurons is different from the plasticity of spatial neurons[2]. The parameters of temporal neurons stay fixed during the simulations, but the neurons can die. As a consequence, what changes in the temporal neural map is the number of surviving neurons. The neuronal death mechanism is inspired from apoptosis (Ameisen, 2000), and fits with the theory of neural epigenesis developed by (Changeux, 1983). The theory basically proposes that neural epigenesis consists of an initial massive generation of random neurons and connections, which are afterwards pruned and selected according to the level of neurotrophins they receive. Neurotrophins are provided to the neurons which are often activated, and prevent them from automatic suicide (Ghosh, 1996). We apply this principle of generation and pruning to our temporal neurons, and depending on their mean

[1] The neural network that we use is technically very similar to Self-Organizing Feature Maps (Kohonen, 1982). In our case, the input space is of the same dimensionality than the output space, so we do not use it to make dimensionality reduction. Feature maps are normally used to extract some regularities in high dimensional input data. Here, there is no regularity in the input data initially. Input data is generated by other neural networks of the same kind. Regularities are rather created through self-organization as explained in the "dynamics" section.

[2] Yet, some recent experiments which we do not describe in this paper because they were not conducted with the same systematicity, indicate that it is possible to use for both neural maps the same neural dynamics and still obtain results similar to those we present. In these experiments, the common neural dynamics was the same as the one we use here for the temporal neural map.

activity level. The mean activity of a temporal neuron j is computed with the formula:

$$MA_{j,t} =$$

$$< GT_{j,t} >= \frac{< GT_{j,t} > .(window - 1) + GT_{j,t}}{window}$$

where $window$ has the initial value 50. The initial value $MA_{j,0}$ is equal to $2.vitalThreshold$. The $vitalThreshold$ constant defines the level of activity below which the neuron is pruned. This threshold remains the same for all neurons in the map. The value of this threshold is chosen so that there is not enough potential activity for all the neurons to stay alive: stability arises at the map level only after a certain amount of neurons have been pruned.

2.3 The coupling of perception and production

The crucial point of this architecture is that the same neural units are used both to perceive and to produce vocalizations, both in the spatial and in the temporal neural map. As a consequence, the distribution of targets which are used for production is the same than the distribution of receptive fields in the spatial neural map, which themselves adapt to inputs in the environment. This implies for example that if an agent hears certain sounds more often than others, he will tend to produce them also more often than others. The same phenomenon applies also to the order of the articulatory targets used in the vocalizations. If an agent hears certain combinations often, then this will increase the mean level of activation of the corresponding temporal neurons, which in turn increases their chance of survival and so increases the probability that they will be used to produce the same articulatory targets combinations. These coupling create positive feed-back loops which are the basis of the self-organization that we will now describe.

It is important to see that this is not realized through imitation, but is a side effect of an increase of the sensitivity of neurons, and of the competition for neurotrophins between the temporal neurons, which is are very generic local low-level neural mechanisms. Agents do not imitate each other in this artificial system, since they never repeat a sound that they just heard, and they never store explicitly a sound that they hear in order to reproduce it later on. Additionally, agents do not play any language game in the sense used in the literature (Steels, 1997). In fact, they have no social skill at all. They are just in a world in which they wander around and sometimes produce sounds and adapt to the sounds they hear around them.

3 The dynamic formation of phonotactics and patterns of combinations

In these simulations, we use a population of 10 agents. As initially the preferred vectors of the spatial neurons are random, and as there is a massive number of random temporal neurons, agents produce vocalizations which are holistic and inarticulate: the continuum of possible articulatory targets is used, and nearly all possible sequences of targets are produced. The initial state of both neural map in two agents is represented on figure 1: the spatial map is represented on the x-axis, which shows the preferred vectors, and is also represented on the y-axis, which shows the same information. The temporal map is represented by the small segments in the middle of the figure, which all correspond to a point (x, y) for which x corresponds to an existing preferred vector in the spatial map, and y to another existing preferred vector in the spatial map. The x coordinate of a temporal neuron corresponds to the first articulatory target of the vocalization that it encodes, and the y coordinate corresponds to the second target that it encodes. The length of the segment represents the level of neurotrophins that each neuron possess.

After several hundred time steps, as we have shown and explained in details in (Oudeyer, in press), we observe a clustering of the preferred vectors of the spatial map. Figure 2 and 3 shows an example of the neural maps after 1000 interactions in two agents (taken randomly among the 10 agents). Moreover, the clusters are the same for all the agents of the same simulation, and different for agents of different simulations. This shows that now the vocalizations that they produce are discrete: the articulatory targets that they use belong to one of several well defined clusters, and so the continuum of possible targets has been discretized.

Moreover, if we observe the temporal map, we discover that there remains only temporal neurons coding for certain articulatory target sequences. This means that some sequences of targets belonging to the spatial clusters are not produced any more. All the agents of the same population share not only the same clusters in the spatial map, but they also share the same surviving temporal neurons, as figures 2 and 3 show. This means that rules of phoneme sequencing have appeared, which are shared by all the popu-

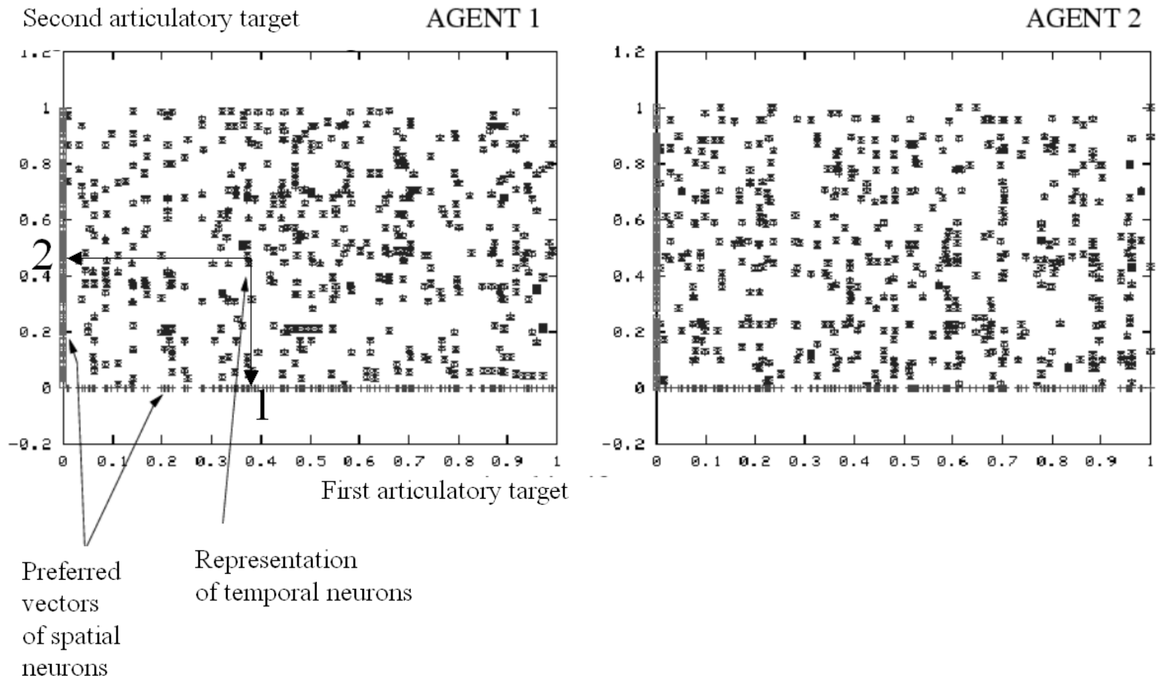


Figure 1: The neural maps of two agents at the beginning of the simulation. The neural maps of one agent is represented on the left, and the neural maps of the other agent are represented on the right. The spatial map is represented by its preferred vectors plotted on the x-axis and also plotted on the y-axis. The temporal neural map is represented by small segments whose centre has its x and y corresponding to preferred vectors of the spatial neural map. The x coordinate of a temporal neuron corresponds to the first target that it encodes, and the y coordinate corresponds to the second target that it encodes.

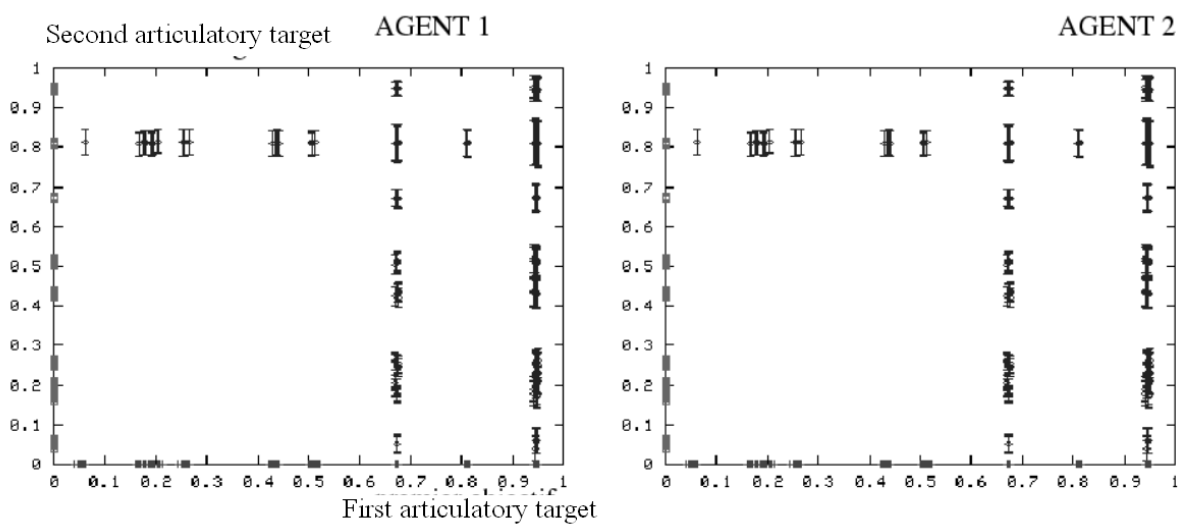


Figure 2: The neural maps of the same two agents after 1000 interactions. We observe: 1) that the preferred vectors of the spatial neural map are now clustered, which means that vocalizations are now discrete: phonemic coding has appeared; 2) that many temporal neurons have died and the surviving ones are organized into lines and columns: this means that phonotactic rules have appeared and moreover that the repertoire of vocalization can be organized into patterns.

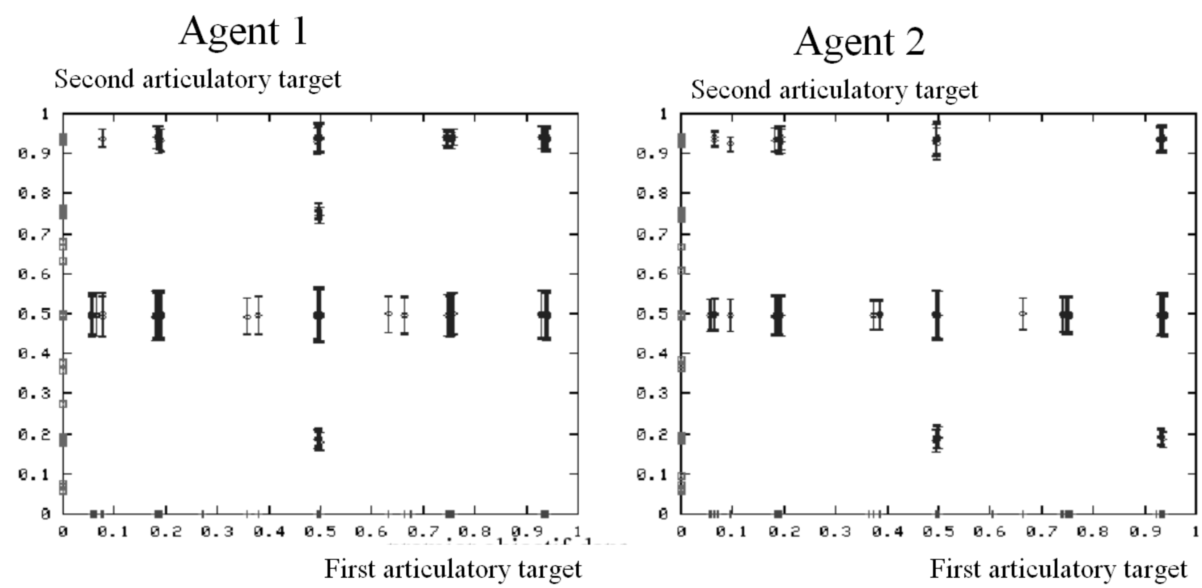


Figure 3: Another example of neural maps of two agents after 1000 interactions.

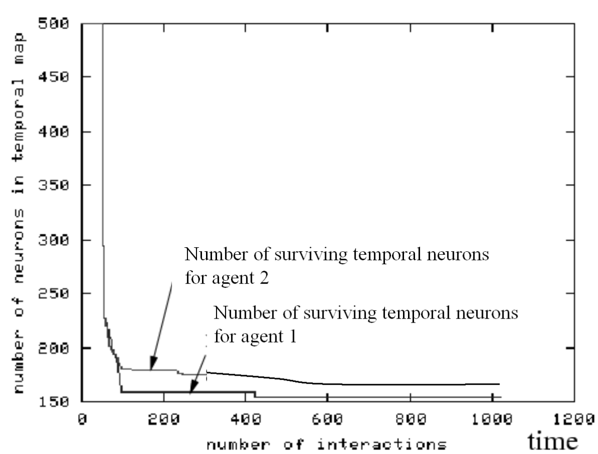


Figure 4: Evolution of the number of surviving temporal neurons corresponding to the temporal neural map of the two agents in figure 2. We observe that there is a first phase of massive pruning, followed by a stabilization which corresponds to a convergence of the system.

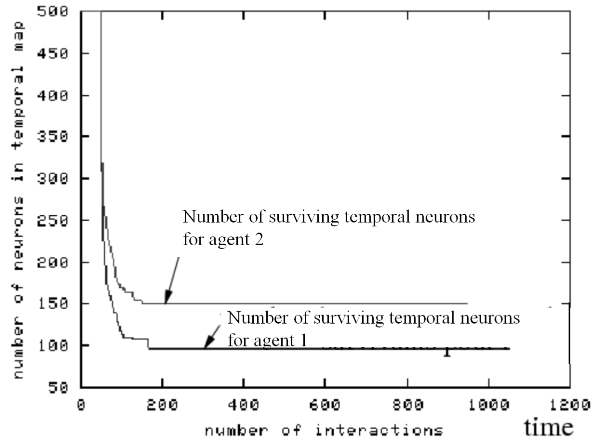


Figure 5: Evolution of the number of surviving temporal neurons corresponding to the temporal neural map of the two agents in figure 3.

lation. In brief, this is the self-organization of phonotactics. Yet, this is not all that we can observe from the temporal neural map. We also see that the surviving temporal neurons are organized into lines and columns. This means that the set of allowed phoneme sequences can be summarized by patterns. If we call the phonemes associated with the eight clusters of the spatial map on figure 2

$$p_1, p_2, ..., p_8$$

then we can summarize the repertoire of allowed sequences by:

$$(p_6, *), (p_8, *), (*, p_7)$$

where $*$ means "any phoneme in $p_1, ..., p_8$". The repertoire is thus organized into patterns, in a manner similar for example to the "CV/CVC/VC" organization of syllables in Japanese.

The states shown on figures 2 and 3 are convergence state. Indeed, both the state of the spatial map and of the neural map crystallize after a certain amount of time. In (Oudeyer, in press), we explained in detail why the spatial map practically converged into a set of clusters for wide range of values of the parameter σ which determines the dynamics of spatial neurons.

We will now explain why there is a convergence in the dynamics of the temporal neural map, as figures 4 and 5 show (we have plotted the evolution of the number of surviving neurons within the temporal maps of two agents). As explained above, the initial level of activity ($MA_{j,0}$) of the temporal neurons is set to a constant ($2.vitalThreshold$) which is higher than the mean level of activity that will be actually computed for each neuron at the beginning of the simulation when they are still all alive. As a consequence, the mean level of activity of all neurons is going to go down at the beginning of a simulation. Because there is stochasticity in the system, due to the random choice of temporal neurons when a vocalization is produced, and also due to the fact that all uniform distributions of preferred vectors are not exactly the same in different agents, all the $MA_{j,t}$'s will not decrease exactly in the same manner. In particular, certain temporal neurons will have their $MA_{j,t}$ go below the vital threshold ($vitalThreshold$) before the others and die (indeed, $vitalThreshold$ is chosen so that it is higher than the mean level of activity of neurons if they are all alive). The survival of one temporal neuron in a cluster of the temporal map of one agent ag depends on the number of neurons in the corresponding cluster in other agents, whose survival depends in return on the number of neurons in the cluster of the agent ag. This creates positive feedback loops: sometimes and by chance, a number of neurons die in the same cluster of one agent, which favours the death of similar neurons in other agents, because having less neurons in one cluster or area of the space decreases the probability to produce a vocalization coded by the neurons of this cluster and so decreases the mean level of activity of the corresponding cluster in the other agents. Reversely, clusters composed of neurons with a high mean level of activity will favour the survival of similar clusters in other agents. This interaction between the competition and the cooperation in the clusters of temporal neurons of all agents will precipitate a number of neurons, and a number of clusters of neurons, below the vital threshold, until there remains few enough clusters so that the neurons that compose them are activated often enough to survive and "live" together. This explains the stabilisation observed on figures 4 and 5, where we see the two phases: a first phase of initial and rapid pruning of neurons, and a second phase of stabilisation.

The "cooperation" / positive re-inforcement can happen between clusters of temporal neurons coding for the same phonemic sequence, but also between clusters of temporal neurons sharing only one phoneme / articulatory target at the same location within the vocalization. This is due to the mode of activation of temporal neurons, as detailed in the formula above. For example, let us denote p_1, p_2, p_3 and p_4 four distinct phonemes / articulatory targets belonging to four distinct clusters. If the similarity of two vocalizations with the same sequence of phonemes is about 1, then the similarity between the vocalization coded by the sequence (p_1, p_2) and the vocalization coded by the sequence (p_1, p_3) is about 0.5, and the similarity between (p_1, p_2) and (p_3, p_4) is about 0. This means that the level of activity "provided" to the temporal neurons of a cluster cl thanks to two clusters of temporal neurons in other agents which share exactly one phoneme in the same location, is about the same that the level of activity provided to the neurons in cl thanks to the cluster in other agents which corresponds to temporal neurons sharing all the phonemes in the right location with those in cl. As a consequence, groups of clusters re-inforcing each other will form during the self-organization of the temporal neurons map. These are the lines and the columns that we observed on figures 2 and 3, and this explains why we observe the formation of phonological patterns in the phonotactics developed by the agents. To summarize, the interactions between competition and cooperation among individual clusters explains the formation of shared and stable repertoires of allowed phoneme sequences, and the interaction between competition and cooperation among groups of clusters explains the formation of phonological patterns.

4 Conclusion

In (Oudeyer, in press), we presented a system showing how a society of agents could self-organize a discrete speech code shared by all speakers of the same community, and different in different communities. We also showed how it allowed to predict certain statistical regularities characterizing the repertoires of phonemes in human languages. The originality of the system was that: 1) it did not include any pressure for communication; 2) it did not include any social capacity (agents did not play a language game for example); 3) it pre-supposed neither linguistic capacities nor the existence of conventions. This made the system a good tool to think and develop our intuitions about the bootstrapping of speech, and attack the problem of the origins of language as opposed to the problem of the formation of languages which has already been studied extensively in the computer modelling literature (e.g. Browman and Goldstein (2000); Kaplan (2001); Kirby (2001); de Boer (2001); Oudeyer (2001); Cangelosi and Parisi (2002)). Indeed, by making evolutionarily simpler assumptions than existing models, one can understand more easily how natural selection, in an environment favouring the reproduction of individual capable of communication, could have been guided by self-organization to establish the first and primitive forms of conventions, such as the speech codes that our agents generate. In this paper, we presented a natural and crucial extension to our earlier work, introducing a mechanism that takes into account the order of articulatory targets both in production and in perception of vocalizations. This allowed to show that similarly, agents could self-organize phonotactics, defining shared sets of allowed phonemic sequences in a given population. Diversity was again a feature: different populations of agents developed different phonotactics systems. Moreover, the set of allowed phonemic sequences could always be organized into patterns, which has strong structural similarities with the phonological patterns that we observe in human languages.

Yet, with the system presented in this paper, we can not show a statistical preference for certain kinds of phonotactics and patterns as compared to others. This is indeed also a feature of human phonological systems: for example, most languages allow CV syllables, but many disallow consonant clusters at the beginning of syllables like in CCV syllables. A possibility to account for this phenomenon is to introduce morpho-perceptual constraints, as it allowed the prediction of human vowel systems in (Oudeyer, in press), and energetical constraints such as the cost of articulation of vocalizations. We are currently working on this particular issue, and the results will be discussed in a forthcoming paper.

References

J.C. Ameisen. *La Sculpture du vivant. Le suicide cellulaire ou la mort cratrice.* Seuil, 2000.

P. Ball. *The self-made tapestry, Pattern formation in nature*. Oxford University Press, 2001.

C.P. Browman and L. Goldstein. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parle*, 5:25–34, 2000.

A. Cangelosi and D. Parisi. *Simulating the evolution of language*. Springer Verlag, 2002.

J.P. Changeux. *L'homme neuronal*. Fayard, 1983.

B. de Boer. *The origins of vowel systems*. Oxford Linguistics. Oxford University Press, 2001.

A. Ghosh. Cortical development: With an eye on neurotrophins. *Current Biology*, 6:130–133, 1996.

F. Kaplan. *La naissance d' une langue chez les robots*. Hermes Science, 2001.

S. Kirby. Spontaneous evolution of linguistic structure - an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

B. Lindblom. Phonological units as adaptive emergents of lexical development. In Ferguson, Menn, and Stoel-Gammon, editors, *Phonological Development: Models, Research, Implications*, pages 565–604. York Press, Timonnium, MD, 1992.

P-Y. Oudeyer. The origins of syllable systems : an operational model. In J. Moore and K. Stenning, editors, *Proceedings of the 23rd Annual Conference of the Cognitive Science society, COGSCI'2001*, pages 744–749. Laurence Erlbaum Associates, 2001.

P-Y. Oudeyer. *L'auto-organisation de la parole*. PhD thesis, Université Paris VI, 2003.

P-Y. Oudeyer. The self-organization of speech sounds. *Journal of Theoretical Biology*, in press.

J.L. Schwartz, L.J. Boe, N. Valle, and C. Abry. The dispersion/focalization theory of vowel systems. *Journal of phonetics*, 25:255–286, 1997.

L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1, 1997.

Selection, domestication, and the emergence of learned communication systems

Graham Ritchie[⋆†]
[⋆]Centre for Intelligent Systems and their Applications
University of Edinburgh
graham.ritchie@ed.ac.uk

Simon Kirby[†]
[†]Language Evolution and Computation Research Unit
University of Edinburgh
simon@ling.ed.ac.uk

Abstract

One of the most distinctive characteristics of human language is the extent to which it relies on learned vocal signals. Communication systems are ubiquitous in the natural world but vocal learning is a comparatively rare evolutionary development (Jarvis, 2004). In this paper we take one example of this phenomena, bird song, which displays some remarkable parallels with human language (Doupe and Kuhl, 1999), and we focus on one particular case study, that of the Bengalese finch (*Lonchura striata* var. *domestica*), a domesticated species whose song behaviour differs strikingly from its feral ancestor in that it has complex syntax and is heavily influenced by early learning (Okanoya, 2002). We present a computational model of the evolutionary history of the Bengalese finch which demonstrates how an increase in song complexity and increased influence from early learning could evolve spontaneously as a result of domestication. We argue that this may provide an insight into how increased reliance on vocal learning could evolve in other communication systems, including human language.

1 Introduction

The human capacity for language is one of our most distinctive characteristics. While communication systems abound in the natural world, human language distinguishes itself in terms of its communicative power, flexibility and complexity. One of the most unusual features of human language, when compared to the communication systems of other species, is the degree to which it involves learning. Just how much of language is innate and how much is learned is an ongoing controversy, but it is undeniable that the specific details of any particular language must be learned anew every generation. We do, of course, bring a great deal of innate resources to bear on our language learning process, and the results these innate biases have on the development of languages may explain a great deal about the structure of the languages we see today. But still every child in every new generation must go through a lengthy process of language acquisition if they are to become normal language users.

Once in place, this inter-generational process of language acquisition and use, or *iterated learning* (Kirby and Hurford, 2002), can give rise to cultural evolution, which studies have shown may explain many prominent phenomena of human language, including the emergence of dialects and, by extension, separate languages (Livingstone, 2002), regular and irregular word forms (Kirby, 2001) and compositional syntax (e.g. Brighton, 2002).

The emergence of learning can therefore be seen as a major transition in the evolution of language and we would like to better understand the evolutionary pressures and factors which caused this transition. A natural point at which to start such an investigation is to look at the communication systems of other animals to see if there are any parallels which might illuminate the relevant ecological factors. Much comparative research has been carried out with the non-human primates, but despite some fascinating results, it seems that their natural communication systems are very different to language, including the fact that learning plays a much less prominent role. In fact it appears that vocal learning systems have evolved in only three groups of mammals: humans, bats and cetaceans, and three groups of birds: songbirds, hummingbirds, and parrots (Jarvis, 2004).

In this paper we concentrate on bird song as it has many striking parallels with language, particularly the way in which it is learned, as Darwin noted in *The Descent of Man*:

> The sounds uttered by birds offer in several respects the nearest analogy to language, for all the members of the same species ut-

> ter the same instinctive cries expressive of their emotions; and all the kinds that sing, exert their power instinctively; but the actual song, and even the call notes, are learnt from their parents or foster-parents. (Darwin, 1879, p. 108)

Since Darwin's day much research has been carried out into bird song and, to take Tinbergen's four perspectives of ethology, we now know a great deal about its mechanism, development, function and evolution. However, despite much research, in general the evolutionary function of song learning remains unclear (Slater, 2003). The parallels between bird song and human language have also been further elaborated as modern techniques have allowed us to establish the neural mechanisms of both song and language (Doupe and Kuhl, 1999).

2 A case study

Recent studies by Kazuo Okanoya of a domesticated species of finch, the Bengalese finch (*Lonchura striata* var. *domestica*), and its feral ancestor, the white-backed munia (*Lonchura striata*), provide an interesting case study of the interaction of learning and evolution in bird song. The Bengalese finch sings a song with complex finite state syntax which is heavily influenced by early auditory experience. Surprisingly, the munia sings a strikingly simpler, more linear song which is less influenced by early learning. In other words, in a relatively short period of domestication, there have been radical changes in song behaviour. This has happened even though the domesticated species has been artificially bred for plumage rather than song.

Okanoya (2004) has identified the neural mechanism underlying this difference in behaviour and has shown that while Bengalese chicks are able to learn the songs of munia tutors, munia chicks are not able to learn the more complex Bengalese song, clearly demonstrating that there is a physiological basis for this difference.

2.1 Okanoya's hypothesis

As experiments have shown that both female munias and female Bengalese finches prefer the more complex song, Okanoya (2002) argues that it is sexual selection which drove this increase in complexity. He argues that domestication freed the Bengalese finch from the pressure of predation and other pressures associated with life in the wild which had previously held song complexity in check. According to Okanoya, the more complex song of the Bengalese finch may therefore be seen as an honest signal of fitness (Zahavi, 1975); a fitter bird can afford a more complex song. Sasahara and Ikegami (2004) show with a computational model of the finch data that song complexity could indeed increase as a result of sexual selection.

2.2 Deacon's hypothesis

Reviewing the same data, Deacon (p.c.) agrees that domestication masked the natural selection pressure keeping the munia's song simple, but argues that the increase in complexity happened *without* direct selection on the trait. Essentially, he posits that domestication shielded the trait from selection which allowed random genetic drift to erode innate song biases in the munia. This allows previously minor influences, such as mnemonic biases and early auditory experience, to have more of an effect on song structure and learning, which results in the various neural modules involved in song production and learning becoming increasingly *de-differentiated*. Deacon goes on to argue that this process of masking and subsequent de-differentiation is a potential explanation for the evolution of complex functional synergies such as the neural mechanisms for song production now present in the Bengalese finch, and, he argues, in the human capacity for language. The concept of selective masking and its effect on the evolution of language are explored in more detail in (Deacon, 2003).

3 A computational model

In order to evaluate Deacon's hypothesis and to try to establish if such behaviour could evolve spontaneously as a result of domestication, we have developed a computational model of the finch data. The model is designed to be reasonably biologically plausible, and also general enough that it could be extended to other species. The model works with an evolving population of agents, or birds, and the main stages in the simulation are listed here, details of each stage are given below:

Birth The bird's song filter is built up from its genotype as described in section 3.1.

Development The bird is exposed to e songs from its environment, and, using its filter, selects t songs from which it will learn (its training set) as described in section 3.2. The bird then uses the learning algorithm described in section 3.3 to learn the song grammar it will use to sing throughout its life.

Adulthood The bird is tested in f fitness trials, as described in section 3.4 to see how many times, using its filter, is can correctly recognise a bird of its own species and how many times it is correctly recognised by a bird of its own species. These values are added to give a bird's fitness score.

Reproduction Parents of the same species are selected probabilistically according to their fitness score and their chromosomes are crossed over using one-point crossover with probability pCO (set to 0.7 for all results provided here), to give a new child. Individual genes are mutated with probability $pMut$ (set to 0.05 for all results provided here). The mutation operator used is the 'Reflect' operator described in (Bullock, 1999).

Death Each bird in the population is sampled s times and the resulting songs are stored for the next generation to learn from. All of the current birds in the population are removed and their children become the new population.

3.1 The song filter

A bird is modelled as having a genetically coded note[1] transition matrix, which specifies a transition probability from each note to every other note in the used in the simulation, including a probability for the first and final notes. The total number of notes is a parameter of the simulation, $numNotes$, but in all results provided here this was set to 8, i.e. the notes from a alphabetically through to h, this value was chosen as it appears to be the number of unique notes identifiable in both the Bengalese finch and munia's songs (Okanoya, 2002, p. 56). The matrix is coded for by a chromosome which has one real valued locus for each entry in the matrix which can vary between 0 and 1. This chromosome will thus have $(numNotes + 1)^2$ loci, the 1 is added to include the transitions at the beginning and end of the song. To construct a matrix from the chromosome we look at each $numNotes + 1$ loci of the chromosome in turn, and normalise the values to give a probability distribution for each row of the matrix. An example matrix, and the chromosome that codes for it is shown in table 3.1. Note that this scheme allows different genotypes to code for the same phenotype.

The transition matrix serves one main purpose; to establish the probability that a given song is one of the bird's own species song. This is done by establishing the average probability of each note transition in the song, as shown in equation 1 which defines the preference a given matrix m_x has for a particular song s_y, in this equation n is the number of note transitions in s_y and $m_x(t_i)$ is the entry in m_x for the ith transition of s_y. For example the preference value the matrix in table 3.1 gives for the song cab, which has the transitions S-c, c-a, a-b and b-E, is $\frac{0.62+0.82+0.89+0.45}{4} =$ 0.695, while the preference for the song acb is 0.043. Note that we always include the transition to the first note and from the last note, so the empty song '' has a single transition S-E, for which this matrix has a preference value of 0.15.

	a	b	c	E
S	0.08	0.15	0.62	0.15
a	0.11	0.89	0.00	0.00
b	0.05	0.10	0.40	0.45
c	0.82	0.09	0.00	0.09

0.1	0.2	0.8	0.2	0.1	0.8	0.0	0.0	0.1	0.2	0.8	0.9	0.9	0.1	0.0	0.1

Table 1: An example note transition matrix and the chromosome that codes for it. The S indicates the start of the song, and the E indicates the end of the song.

$$preference(m_x, s_y) = \frac{\sum_{i=0}^{n} m_x(t_i)}{n} \quad (1)$$

The matrix can be thought of as a song 'filter'. A song with a high probability will be more likely to pass though the filter than one with a lower probability, in our example cab would be much more likely to pass through the filter than acb. If the matrix has a single high probability transition for each note this can be thought of as a strong filter, as it will only accept songs which contain these transitions. If the matrix has even probabilities for each transition it is considered a weak filter as it accepts all songs equally.

We can measure the strength of the filter explicitly by calculating the entropy for each transition distribution (i.e. each row in the matrix), using Shannon's (1948) measure. This will result in a value which ranges from 0 to $log(nValues)$, where $nValues$ is the number of probabilities in row r_x (i.e. the number of columns in the matrix). We then normalise this value into the range 0 to 1, as shown in in equation 2, which defines the normalised entropy for a given row r_x, in this equation p_i is the probability of the ith transition in r_x. The overall strength of a matrix m_x is then calculated as the average entropy

[1] It should be noted that while we use the term 'note' throughout this paper, this is not intended to refer to a particular acoustic note, rather we simply use it to denote an atomic song element that can be reliably differentiated from other elements which appear in the song.

of each row r in the matrix, as shown in equation 3. A filter strength of 0 means that the filter will only accept one song while an strength of 1 means that the filter will accept all songs equally. As an example, the matrix in table 3.1 has a strength value of 0.56.

$$entropy(r_x) = \frac{-\sum_{i=0}^{nValues} p_i \; log \; (p_i)}{log(nValues)} \quad (2)$$

$$strength(m_x) = \frac{\sum_{i=0}^{nRows} entropy(r_i)}{nRows} \quad (3)$$

This filter is intended to model the preferences many songbirds have for their species specific song (Catchpole and Slater, 1995). In the model a bird uses its filter for two purposes:

1. To select its training set (the songs it will later use to learn from) from the songs it is exposed to during infancy.

2. To judge whether another bird is a member of the same species for mating or territorial defense.

In this respect, this model is similar to those used in Lachlan's models of the 'cultural trap' in bird song (Lachlan and Slater, 1999; Lachlan and Feldman, 2003). This seems a reasonably plausible assumption, as it is known that some songbirds do have an innate preference for conspecific song both when learning songs as a nestling and also for later mate selection (Catchpole and Slater, 1995).

3.2 Selecting the training set

The infant bird is exposed to e environmental songs to select its t training songs from, both e and t are parameters of the simulation, but were set to 50 and 5 respectively for all results provided here. 5 seems a rather low value of t, but the learning algorithm is very computationally intensive and so a low value is used to speed up the simulation. The e environmental songs are randomly selected from the songs sampled from the previous generation, to compose this set each bird is sampled s times, another parameter which is set to 5 here, so for a population size $popSize$ of 100, as used here, this will contain 500 songs.

The infant bird is exposed to each of the e songs in turn and uses its filter to compute the probability it will be accepted. During experimental runs it was determined that checking that the song is accepted once did not impose enough of a pressure for the bird to correctly select conspecific song and so a song is only added if it is accepted by the filter twice successively. If the bird has not picked t songs after being exposed to all e songs, the process is repeated until t songs have been selected. The training songs are then fed into the learning algorithm described below.

3.3 Song learning

Song learning is modelled as minimum description length (MDL) induction of a probabilistic finite state machine (PFSM), closely following the algorithm described in (Teal and Taylor, 2000). Induction of finite-state machines was chosen to model learning as Okanoya (2002) argues that the songs of both munias and Bengalese finches can be usefully described by a finite-state syntax. The algorithm works by firstly establishing the maximal PFSM that explicitly represents each song in the training set, the prefix tree. The algorithm then searches for nodes which can be merged which will reduce the MDL of the overall machine, whilst also ensuring that the PFSM remains deterministic. The MDL measure takes into account the amount of information (measured by the number of bits) required to code for the machine itself, and also to code for each of the training songs in terms of the machine. Essentially the algorithm searches for the most parsimonious machine in terms of the data. This approach allows us the bird to generalise from its training set, whilst also always being able to reproduce each of the songs it learned from. The reader is referred to (Teal and Taylor, 2000) for a more detailed description of the algorithm used. The only difference between Teal and Taylor's and our approach is that we also take into account the probability of each note transition, given the probabilities of each transition in the training set.

3.4 Calculating a bird's fitness

To establish a bird's fitness we want to check both that its filter allows it to correctly identify its own species, and that its song is correctly identified by other birds of its species. This seems a reasonable model of the pressures acting on song in the wild (Catchpole and Slater, 1995).

To calculate an individual bird b_1's fitness we perform f fitness trials, a parameter set to 250 for the results provided here. In each fitness trail we get the b_1 to produce a song and we then randomly select another member of the population, b_2 and check that b_2 correctly recognises the song using its filter. We also get b_2 to produce a song and check that b_1 correctly recognises the song with its filter. Every correct recognition means that b_1's fitness is incremented by 1. With f set to 250, this means that the maximum fitness achievable is 500, or generally $2f$.

3.5 Modelling the finch data

This is a fairly general model of bird song, and so we need set it up to match the data available on the Bengalese finch and the munia as closely as possible. The simulation passes through 3 main phases, each of which runs for 500 generations. The phases are described below.

Phase 1 We know that the white-backed munia has a very stereotyped song and that it seems to only be able to learn songs that match its species-specific song fairly closely (a munia cross-fostered with Bengalese parents is not able to learn its tutor's song). In our model this corresponds to the munia having a strong filter. To simulate this state we seed the environmental songs with a single song type, e.g. $abcdef$. We then run the simulation for 500 generations using the fitness function and learning algorithm described above. As the environment songs are entirely identical the songs that any bird will learn from are always the same, and so they will always induce the same PFSM. This is not meant to be biologically plausible, we simply want the population to develop strong filters for a particular simple song type.

Phase 2 At the end of phase 1 we have a population of birds who sing a stereotypical song and produce offspring with a strong genetic bias to learn that song. To test if the filter can indeed help young birds recognise the appropriate song to learn from in the second phase of the run we start introducing random songs into the bird's environment, this is intended to model hetero-specific song in the environment. We model this by replacing 10% of the s sampled songs with randomly generated songs which use the same notes as the current population and which are constrained to within the same length as the munia songs.

Phase 3 We model domestication of the population simply by ceasing to calculate fitness, but we continue to perform the crossover and mutation operations. The seems a reasonable model of domestication, as in captivity the birds no longer have to recognise their own species to successfully mate or defend their territory as the mating is now controlled by humans and they are kept in aviaries. Domestication can thus be seen to *mask* the selection pressure on these functions. We continue to introduce 10% of random songs into the environment each generation, as it seems a reasonable asumption that the birds will still be exposed to hetero-specific song, or at least other extraneous sounds, in captivity. Experimental results of this setup are described in the next section.

4 Results

The graph in figure 1 shows several measures taken over the course of each of the three phases described above.

The first measure, in red, is how much variation there is, on average, in an individuals repertoire. This is calculated simply as the total edit (or Levenshtein) distance between a number of an individual's songs (set to 10 for all runs shown here). The second measure, in green, is the average filter strength of the population, calculated as described in equation 3. The third measure, in blue, is the average fitness of the population. We do not calculate this measure in phase 3 and so it does not appear for this phase. The fourth measure we include, in blue, is the average grammar encoding length (GEL) of the population's PFSMs, this is a measure of the size, in bits, it would take to encode a PFSM using the measure defined in (Teal and Taylor, 2000). The fifth measure, in brown, is the average song linearity of the songs tested in measure 1, defined as the number of unique notes in each song divided by the number of unique note to note transitions. The sixth and final measure, in yellow, is the average linearity of the population's PFSMs calculated simply as the number of states divided by the number of transitions. A completely linear PFSM would thus have a linearity of 1, while a maximally non-linear PFSM would have a linearity equal to 1 over the number of transitions in the PFSM.

Two example PFSMs taken from the population at the end of phase 2 are shown in figure 2, and two PFSMs from the end of phase 3 are shown figure 3. The GEL and PFSM linearity values for each machine is also given.

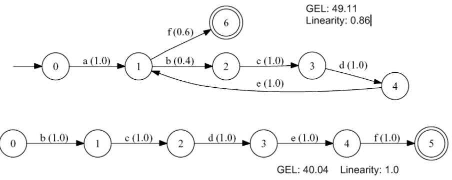


Figure 2: Two example PFSMs from the population at the end of phase 2.

4.1 Analysis

These results demonstrate that the strong filters built up in phase 1, as shown by the increase in filter

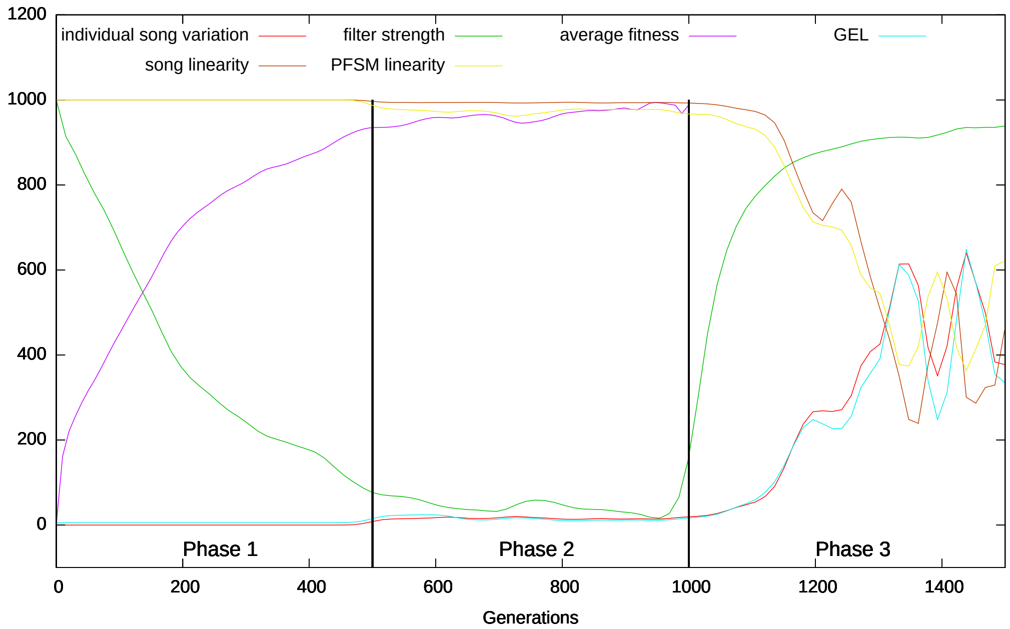


Figure 1: These results are averages taken over 10 separate runs of the simulation with a different random number generator seed for each run. It should be noted that all these measures have been normalised and linearly mapped to give a value within 0 and 1000, where the minimum value seen in the run is set to 0 and maximum values seen in the run is set to 1000. This graph therefore only shows the relative change each of the measures over the course of a run, not the absolute values of each measure. We have also smoothed the lines in the graph to better allow us to see the overall trends. More detailed results are available upon request.

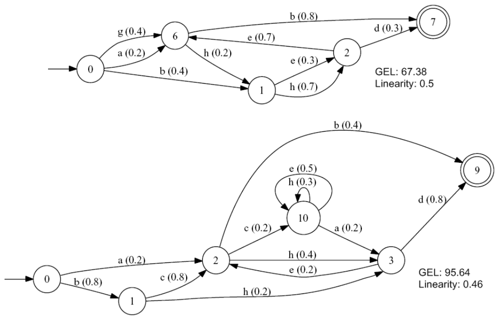


Figure 3: Two example PFSMs from the population at the end of phase 3.

strength[2], enable the birds to filter out the hetero-specific songs introduced in phase 2 without any fitness decrease. We see that all 5 measures stay roughly the same throughout this phase, indicating that this is a fairly stable state. When we 'domesticate' the population in phase 3 we see a significant change in its behaviour. Immediately we see that the filters begin to weaken, and we see that the GEL and the individual variation measures also rise steadily throughout the phase indicating that the population's PFSMs are getting larger and the birds have a more varied song repertoire. At the same time we see both the song, and underlying PFSM linearity drop, indicating that the songs a bird will sing have comparatively more varied note transitions.

This behaviour seems to be a result of the fact the strength of the population's filters is no longer being selectively maintained, that it they have been *masked* from selection. This allows mutations to accumulate and for the filters to become steadily weaker. This allows some of the hetero-specific songs to pass though the filter when a bird is selecting its training set, which results in the bird inducing a more varied PFSM. Essentially the domesticated populations is able to learn from much more varied sources and so early auditory experience has much more of an effect on adult song behaviour.

These results are comparable to those provided in (Wiles et al., 2002), who show with a computational model how a similar masking effect, in this case a fruit-rich environment rather than domestication, may have played a role in the loss of the ability of anthropoid primates to synthesize vitamin C. Their

[2]Note that the strongest filter would give a value of 0, and the weakest 1

model, however, goes further than ours and shows that if the selection pressure were later *un*masked this could result in selection for other abilities, e.g. colour vision, that maintained the levels of vitamin C available.

4.2 Song complexity?

Okanoya (2002) argues that the Bengalese finch has a much more “complex” song that the munia. His measure of complexity is the song linearity, defined as the total number of unique song notes divided by the number of unique note-to-note transitions. He finds that the average song linearity of the munia is around 0.8 while the Bengalese finch song has a value of around 0.4 (p. 56). We provide results for this measure over the course of our simulations in the graph above, but on average we also see a higher value, around 0.95, for the ancestral population and a lower value, around 0.6, for the domesticated population.

While this measure seems a reasonably intuitive measure of song complexity, it should be noted that this measure will classify an entirely random song as maximally complex. We do not want to equate randomness with complexity, but we find it hard to define a measure that can differentiate between the two. Any standard measure of the information content of a song will not be able to do so; a random song is maximally informative in information-theoretic terms. However we consider that two measures, the GEL of a bird’s PFSM taken together with the linearity of the PFSM provide a reasonable estimate of the complexity of a song. A PFSM with a very small GEL and a low linearity is likely to produce more random songs, as it approaches a one state PFSM with multiple transitions back to the same state. A PFSM with a large GEL, but a very high linearity (as we see in the ancestral population in the model) will produce an entirely linear song. A PFSM with a large GEL and a relatively low linearity will produce songs that we are more happy to refer to as complex, as the GEL indicates that it has many states, and so different notes will be used in different contexts, but each state also has several transitions which means that different transitions can be made from each context. Our results demonstrate that the domesticated population does have a higher GEL and a lower PFSM linearity than the wild population and so we are tentatively happy to agree that domestication has caused an increase in song complexity. However, we are still working on developing a more satisfactory measure of song complexity.

4.3 Comparison with the biological data

Comparing these results with the data available for the Bengalese finch we find that the model does seem to capture some of the phenomena involved. Okanoya has shown that a munia chick which is not exposed to conspecific song will not sing a normal song, which seems to fit with the model. He has also shown that while Bengalese chicks can readily learn munia songs, munia chicks cannot learn the more complex Bengalese songs. In the model this difference is attributable to their different filters. The difference in the values for the song linearity in the ancestral and domesticated populations also seem to match fairly well.

As it stands though, the model does not explain why the female munia prefers the Bengalese song. We would argue that a bias for complexity song may have been latent in the munia, and the fact that the munia females prefer the more complex song does not prove that this was the driving force for the change in song behaviour, although introducing such a preference into the model may help to tease these pressures apart. Okanoya (2004) demonstrates that the NIf region of the Bengalese finch’s brain is necessary for the it to be able to sing the more complex song; when surgically lesioned a Bengalese finch with previously complex song will sing a simpler, more munia-like song. We would argue that the model remains neutral to this datum, as it is possible that the munia does have this pathway present in its brain but, because it only ever learns a simpler song, does not use it.

5 Discussion

Our results demonstrate that an increase in song complexity (in some sense) and increased influence from early learning can arise *without* direct selection on either trait, simply through the process of domestication, but what is the significance of this result for the study of human language? Can studying the evolution of learning and complexity in bird song inform our study of the origins of complex language in our species? We believe that understanding the mechanisms behind the emergence of the Bengalese finch’s song, and indeed the evolution of bird song in general, is valuable for evolutionary linguistics in two ways.

Firstly, it has been argued that iterated learning is a key mechanism for the origins of syntax in human language (Kirby and Hurford, 2002). It is striking that human language differs from most other communication systems both in being transmitted through iterated learning and in having complex syntactic structure. We say "most" here but not "all". We appear to be in a very exclusive club with songbirds as another member. Of course, there are important differences between iterated learning in humans and birds. For example, in the former a central constraint on transmitted languages is that they be *expressive*, in that strings must convey complex meanings. Bird song does not carry meaning in the same way, although a diversity of songs may play a role as a sexual display (Catchpole and Slater, 1995). Nevertheless the co-occurrence of iterated learning and signal complexity in both songbirds and humans combined with the rarity of either anywhere else in nature cannot be ignored.

Secondly, and more specifically, by uncovering the crucial role of selective *masking* in the case of the Bengalese Finch, we bring a new mechanism to the table for discussions of the origins of human syntax. It is quite possible that we should not be looking for selective advantages of a culturally transmitted syntactic language, but rather asking what selective forces may have been shielded in our recent evolutionary past. The lifting of selection pressure, and the subsequent diversification of behaviour could have been the necessary precursors of a system of iterated learning for language. What remains to be understood is exactly what more is required for any subsequent modification and synergistic reorganisation of the neural mechanisms underlying these new behaviours.

We feel that the answer to this question is best pursued through computational modelling of the vocal behaviour of both birds and humans.

Acknowledgements

We would like to thank Kazuo Okanoya and Terry Deacon for very useful discussion on this work. We would also like to thank the members of the Language Evolution and Computation Research Unit at the University of Edinburgh for helpful comments and feedback. The first author is supported by an EPSRC studentship.

References

Henry Brighton. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54, 2002.

Seth Bullock. Are artificial mutation biases unnatural? In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Fifth European Conference on Artificial Life (ECAL99)*, pages 64–73. Springer-Verlag, 1999.

C. K. Catchpole and P. J. B. Slater. *Bird Song: Biological themes and variations*. Cambridge University Press, 1995.

Charles Darwin. *The Descent of Man, and Selection in Relation to Sex*. John Murray, London, 2nd edition, 1879. Reprinted in 2004 by Penguin.

Terrence Deacon. Multilevel selection in a complex adaptive system: the problem of language origins. In Bruce Weber and David Depew, editors, *Evolution and Learning: the Baldwin Effect Reconsidered*, pages 81–106. MIT Press, Cambridger, MA, 2003.

Allison J. Doupe and Patricia K. Kuhl. Birdsong and human speech: Common themes and mechanisms. *Annual Reviews of Neuroscience*, 22:567–631, 1999.

Erich D. Jarvis. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016:749–777, 2004.

Simon Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5(2):102–110, 2001.

Simon Kirby and James R. Hurford. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language*. Springer Verlag, London, 2002.

R. F. Lachlan and M. W. Feldman. Evolution of cultural communication systems: the coevolution of cultural signals and genes encoding learning preferences. *Journal of Evolutionary Biology*, 16: 1084–1095, 2003.

Robert F. Lachlan and Peter J. B. Slater. The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. *Proceedings of the Royal Society of London. B*, 266:701–706, 1999.

Daniel Livingstone. The evolution of dialect diversity. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, chapter 5, pages 99–118. Springer Verlag, London, 2002.

Kazuo Okanoya. Sexual display as a syntactic vehicle: The evolution of syntax in birdsong and human language through sexual selection. In Alison Wray, editor, *The Transition to Language*, chapter 3. Oxford University Press, Oxford, 2002.

Kazuo Okanoya. The bengalese finch: A window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences*, 1016:724–735, 2004.

Kazutoshi Sasahara and Takashi Ikegami. Song grammars as complex sexual displays. In *Artificial Life 9*, 2004.

C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656, 1948.

P. J. B. Slater. Fifty years of bird song research: a case study in animal behaviour. *Animal Behaviour*, 65: 633–639, 2003.

Tracy K. Teal and Charles E. Taylor. Effects of compression on language evolution. *Artificial Life*, 6: 129–143, 2000.

Janet Wiles, James Watson, Bradley Tonkes, and Terrence Deacon. Strange loops in learning and evolution, 2002. available from: `citeseer.ist.psu.edu/wiles02strange.html`.

A. Zahavi. Mate selection – a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214, 1975.

Emergence of a Communication System: International Sign

Rachel Rosenstock, Ph.D.
Linguistics Department
Gallaudet University
800 Florida Avenue NE
Washington DC 20002
rachel_rosenstock@yahoo.com

Abstract

International Sign (henceforth IS) is a communication system that is used widely in the international Deaf Community. The present study is one of the first to research extensively the origin of both the IS lexicon and grammatical structures. Findings demonstrate that IS is both influenced by natural sign languages (henceforth SLs) and relies heavily on iconic, universal structures. This paper shows that IS continues to develop from a simplistic iconic system into a conventionalized system with increasingly complex rules.

1 Introduction

The emergence of IS in its present form began in the late 1950s, when the World Federation of the Deaf (WFD) recognized the need for a reliable and structured communication system for deaf people at international gatherings and events. While interpersonal communication seemed to allow for the spontaneous creation of a system of gestures, signs and pantomime, conference settings such as WFD meetings, committee discussions and sports event planning boards required a more elaborate system. A committee was established that consisted of five deaf people from various European countries and the US (Moody 1989). As a result of their work, 1200 signs were published under the name 'GESTUNO'. (The British Deaf Association 1975). From this collection of signs, a partially conventional sign system has evolved which is used today at international conferences as a both mode of presentation and interpretation. IS interpreters stand at international conferences and provide the audience with an "interpretation" of what is said in a natural language (signed or spoken) at the lecture. How is this possible with such a limited number of conventionalized signs and no grammatical structures? How do IS interpreters make lucid arguments, explanations and create abstract propositions? How much of these interpretations is comprehensible to the audience? If this system is not completely iconic, where do the structures come from? Is the system consistent and reproducible or entirely spontaneous? To distinguish the early form of GESTUNO from the more elaborate system used today, the term IS will be used, following Moody (2002). The present study describes both the historical development of IS, as well as the different sources of IS vocabulary and grammar that shed light onto the emergence of IS.

2 History of IS

To the present day, the hearing community still harbors the misconception that sign language is universal. The desire for international communication, or so many believe, found its answer in the "invention" of signed languages or in the inherent iconicity of the system. Even deaf people believed until the beginning of the 20th century that sign language was an ideal means of communication across borders. At what might have been the first international gathering of deaf people in 1850, the French deaf educator Ferdinand Berthier observed seeing various deaf people from all over Europe interacting with ease. He remarked that '[f]or centuries, scholars from every country have searched for the universal language and failed. Well, it exists all around you, and it is Sign Language!!' (quoted in Moody 2002:10).

In the early 1900s, deaf people started to recognize that communication with gestures alone did not suffice for the increasingly more sophisticated exchanges required on an international level (see Moody 2002). A proposal to form a committee on the creation of an international sign system was introduced by a Finnish deaf delegate at a conference for deaf education in 1911. It was accepted but never actualized (ibid:12). It was at the same conference, Moody believes, that international signs were, at least to a degree conventionalized. \ The need for international communication arises out of the necessity to exchange ideas. In the Deaf Community, with the founding of the WFD and CISS (International Committee of Sports for the Deaf), this necessity grew to a degree that the loosely established 'international' signs did not suffice for satisfying communication needs. It was not until 1977 that the WFD published their compilation of 1200 international signs, called Gestuno (The British Association of the Deaf 1975). Moody (see above) summarizes the process of developing the dictionary and the reactions by the Deaf Community following the publication: "The task was enormous, given the highly flexible and uncoded nature of International Sign, the uncontrolled natural evolution it had followed since the beginning of the 19th century, and the logistical problems involved in calling meetings of the commission (in spite of the fact that the members were all European and American). […] Deaf people soon began complaining that the signs in the Gestuno lexicon weren't iconic enough to be readily understood." (Moody 2002:16).

3 Evolution

Early literature on GESTUNO suggests that the Deaf Community presumed that a collection of 1200 signs would provide enough basis for a full-fledged communication system. This is expressed in the first publication of the GESTUNO dictionary: '[This is a] sign language which would be of assistance to deaf brethren throughout the world enabling them to understand each other at friendly gatherings and official conferences.' (The British Deaf Association 1975, Preface). These signs included in the dictionary were either loan signs from the national SLs used by the committee members, or highly iconic in nature. The assumption was that this collection was easy to learn by both the presenters or interpreters who would use GESTUNO, as well as by the audience at international gatherings. Reports by the first consumers of GESTUNO at an international conference in Bulgaria in 1976, however, are a testimony to the impossibility of expecting an adequate translation by providing interpreters with only a limited list of vocabulary. Moody (2002) describes how the hearing Bulgarian interpreters use spoken Bulgarian grammatical structures, combining them with the GESTUNO signs and Bulgarian SL when GESTUNO signs are not available. The audience reports that they cannot follow any of the presentations or interpretations in GESTUNO. In the following years, native signers, both hearing and deaf, were taking over interpreting at international events.

Moody observes that the vocabulary of IS interpreters can be traced to three main sources: borrowed signs of the native languages of participants of a given discourse, pantomime, and a more conventionalized pool of vocabulary that might vary from group to group but is perceived by its users to be understood universally (1989:94). Moody emphasizes the influence of the sign language of the host country of any given event on the vocabulary of IS there. This is empirically supported by a study on the origin of IS vocabulary conducted by Woll (1990). Her data was collected exclusively in the UK. Over 70% of all IS signs were labelled borrowings from BSL. Other IS signs seem less conventional. Moody (2002) describes them as 'mimed actions'; Locker McKee and Napier describe them as "more iconic, simpler in form" (2002:48). This lack of conventionalized vocabulary could be explained by the lack of native speakers and the use of IS in only a limited number of domains. The role of the invented signs published under the name 'GESTUNO' is not discussed in recent articles. Woll (1990) mentions 13-21% signs that are nonce signs (signs that are created only for the purpose of a specific context) and calls these international signs. Whether or not she includes GESTUNO signs in this category is unclear.

The source of IS grammar, on the other hand, is still widely debated. The few empirical studies of IS in recent years take very different perspectives on the issue. The ability of the interpreters to string together words, express relations between people and objects, and convey meaning beyond the isolated sign indicates that a grammatical system exists. Allsop et al. (1994) summarize the different proposed sources for the grammar of IS as follows:

- "Signers export a fairly complete version of their own Sign Language grammar (Webb/Supalla [1994, in Allsop et al. 1994]);
- Signers export those parts of their own Sign Language grammar 'felt' to be most universal (Moody [1979, in Allsop et al. 1994]);
- Signers use a grammar specifically belonging to International Sign (what Garretson calls 'natural order' [1990, in Allsop et al. 1994]) which is different from the grammar of their own Sign Language;
- Signer uses some combination of their own Sign Language grammar and some structures particular to International Sign"

Moody (1989) notes that the grammatical properties of natural sign languages are important in the formation of IS. But while Supalla and Webb (1994) claim that the entire grammar is included in the formation of IS, Moody singles out only the most salient and universal structures. "Signed phrases use space, modifications of movements of signs and grammatical facial expressions that have been described by linguists for different sign languages and that seem to be common among most or all sign languages." (1989:90). Garretson (1990) does not recognize the influence the native sign language of the user might have on the structure of his/her particular IS. He describes the communication system as a "zone of 'pure' or non-language stripped of grammar and artificial … rules, the only syntax being one of natural order." (44). Allsop et al. (1994) take a more moderate stand among these extreme views. Their notion is that the signer uses a combination of the grammar of their own sign language and structures that are particular for IS. In the present study, the make-up of IS vocabulary and structure as it is used today will be compared to possible sources.

4 Data

Three IS interpreter teams consisting of two interpreters each were filmed for this study. All participants were recruited during an international Deaf Community event and were filmed in their natural work environment. The teams were approached based on the topics of the lectures and the linguistic background of the presenters. Having worked at the event for three days prior to this taping, it can be assumed that the IS interpreters were more familiar with the vocabulary and more comfortable with the structures of IS than on the first day of the event. All participants are native signers of either British SL or American SL, except for one who acquired American SL later in life. Three are also fluent in Australian SL. Other information collected on the IS interpreters includes certification, experience as an interpreter, mode of acquisition of IS and frequency of use.

Of the six 45-minute videotapes of IS interpretation collected at the event, twelve 5-minute clips were randomly chosen and transcribed. The topics of these clips varied from strongly deaf-related topics such as deaf-blindness to comparisons of the economic situation for deaf entrepreneurs in developed and developing countries. All clips were glossed. The glosses were entered into an excel spreadsheet. Various factors were then transcribed. Handshapes for each sign were recorded and counted. Occurrences of fingerspelling, numbers, proper nouns, different types of verbs were counted. Grammatical facial expressions and other non-manual structures were also noted in the spreadsheet.

In the present study, an attempt was made to determine sources of the IS lexicon by a comparison to a variety of natural sign languages. The number of occurrences for each sign in the transcribed data was counted. A list of signs occurring five or more times was compiled. For a determination of the origin of these 162 signs, native signers of 15 SLs were consulted. They were asked to translate all 162 into their native SLs. These forms were compared with the IS form used by the interpreters. In order to determine the degree of universality of signs, the SLs included in this comparison were divided into different groups (henceforth language groups). The division was derived on the basis of descriptions of historical relations between SLs in Woll et al. (2001:25ff) as well as observations made in this study during the review of vocabulary provided by the informants. SLs were categorized into one of five groups: European SLs, ASL, BSL, Asian SLs, and Near Eastern SLs. Since the ASL, BSL and European SLs groups are historically strongly related, those three occurrences were grouped in one category (Western SLs). Asian and Middle Eastern SLs were grouped in another category (Eastern SLs). The categorizations made for the purpose of this study are shown below.

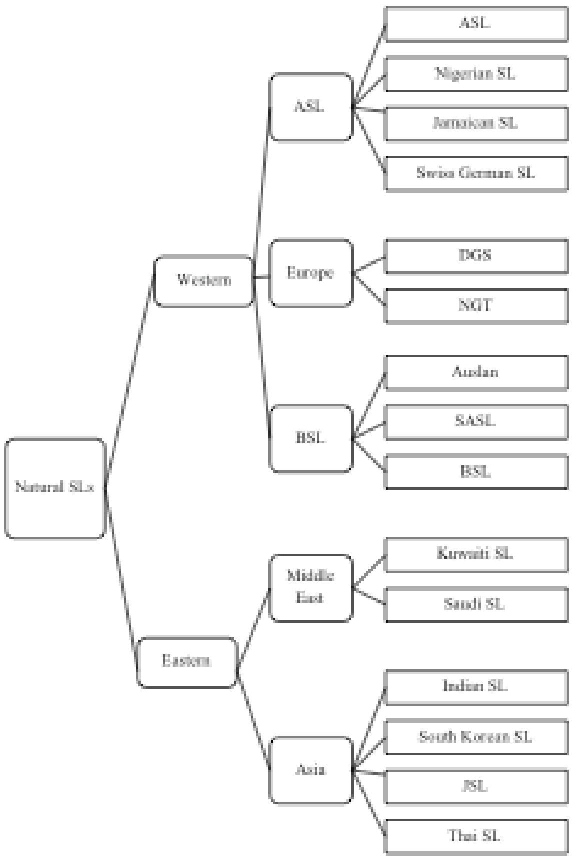


Figure 1: Language Groups

5 IS Vocabulary

As described above, the source of the IS lexicon has been the subject of many discussions. While the general assumption that the vocabulary is based mostly on GESTUNO is easy to disprove (see below), the origin of most IS signs has been said to stem from the local SL where IS is used. Moody (2002) suggests that IS interpreters intentionally adapt vocabulary from the local SL in the assumption that the audience will be most familiar with that particular set of signs. Woll (1990) finds that in her data of IS used in Great Britain, 70 to 80% of the signs were based on BSL. In Woll's study, interpreters also created nonce signs for the specific context of use, employing metonymic and metaphoric expressions to convey the content.

After determining these categories, the use of different form/meaning pairs was counted in all SLs. For example, the IS sign DIFFERENT, shown below, is used in identical form with the same meaning in eight different SLs (namely BSL, NGT, DGS, Thai SL, Swiss-German SL, Jamaican SL, ASL and Nigerian SL). Based on the categorization described above, the IS sign DIFFERENT occurs in the same form in four out of the five SL groups.

Figure 2: IS sign DIFFERENT

Lastly, signs that occurred in three or more unrelated language groups were labeled 'common'. Within this group, those signs occurring in over four of the five language groups and also in over 10 individual sign languages were labeled 'very common'. Examples are the IS signs HOUSE or OLYMPICS, shown below.

In this limited set of data (162 signs), most signs were labeled 'common'. Over 60% were found in three or more unrelated language groups. Only 2% were identified as unique to IS. The remaining 36% were identified as loans from specific SLs or Sign Language groups. Original GESTUNO signs did occur in the current IS data. All GESTUNO signs, however, were iconic to a degree that they were counted into the category labeled 'common'.

5.1 Gestuno

A comparison with the GESTUNO signs shows that most of the original signs were replaced by either more iconic signs or other loan vocabulary. Comparisons of the old GESTUNO signs and their new IS counterparts are shown below. GESTUNO signs are above the gloss; the row beneath shows the forms used today.

Figure 3: Comparison of GESTUNO and IS Signs

5.2 Signs Unique to IS

Only 2% percent of the signs included in this comparison were not found in natural SLs and thus labeled unique to IS. It is possible that these signs do occur in other natural SLs not included in this study.

5.3 Loans

Loans make up a significant part of the IS data. As described above, different categories were distinguished in order to examine the degree of distribution of a sign. 12% of the IS signs compared in this study are loans from a single SL. 15% are loans that occur multiple times in SLs of the same language group. In 20% of the cases, the sign can be found in one Western and one Eastern SL. These are labeled ‘Two Families’. The largest group of signs (53%) are shared by several of the Western SL families. Note that no loans are shared by Eastern families alone. This supports the notion that IS is based largely on Western SL structures.

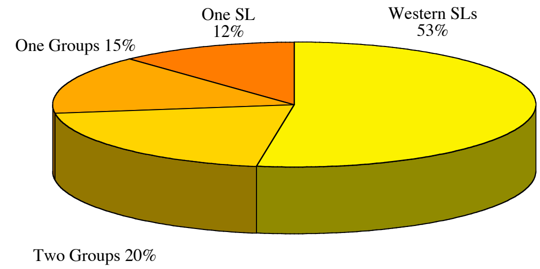


Figure 4: Sources of IS Loans

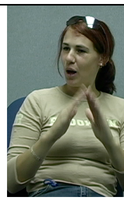
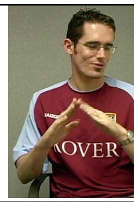
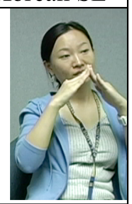
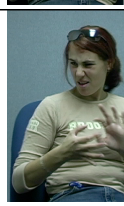
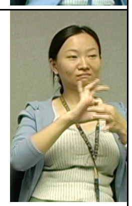
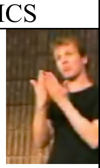
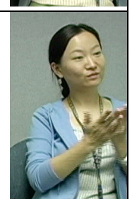
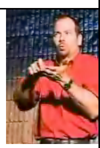
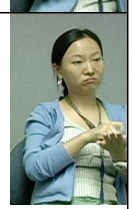
Most loans can be traced back to several SLs within one or two language groups. In the figure below, the top row shows the IS signs for HAVE and GOVERNMENT. The screenshots below show the sign in a natural SL. HAVE is unique to the ASL language group. It occurs in ASL, Jamaican SL and Swiss German SL. GOVERNMENT is found in the BSL group, occurring in both Auslan and SASL.

Figure 5: IS Loan Vocabulary

5.4 Common Signs

This is the largest category of IS signs in this set of data. The most common signs, occurring in over 10 different SLs from four or five of the language groups, account for 24% of the total number of signs included in this study. 38% of the 162 signs are shared by three to five of the language groups in both Western and Eastern SLs. In the signs most common among all SLs, the degree of iconicity is striking. The figure below shows the signs for HOUSE, OLYMPICS, BOOK, WRITE in three SLs from different language groups. The IS sign is shown on the left, followed by the examples from different SLs.

Table 1: Common Signs in IS

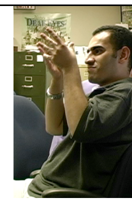
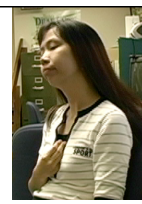
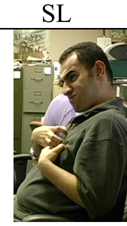
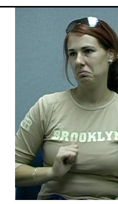
IS Sign	Dutch SL	Australian SL	South Korean SL
 HOUSE			
 OLYMPICS			
 BOOK			
 WRITE			

5.5 Iconicity

As can be seen in the comparison of different sign languages above, iconicity plays an important role in the lexicon of IS. Woll (1990) has labeled 20-30% of all IS signs iconic or metaphoric in nature. The degree of iconic transparency a sign has for an IS consumer is highly dependent on the degree of shared cultural experience of the signer and addressee.

The signs used in the informants' natural SLs all use the same metaphors for knowledge and emotion. The location of the head represents knowledge; the chest has metaphoric meaning representing feeling (described for ASL in Taub 2001 and for BSL in Brennan 1990). Signs for the concept of 'increasing knowledge' are shown in the upper row. A depressed feeling is expressed in the signs in the row below. A non-signer was able to guess meanings related to knowledge and feeling.

Table 2: Universal Metaphors 'feel'/'know'

			
Thai SL	Kuwaiti SL	NGT	SKSL
			

IS makes use of the universality of these metaphors in several signs in combination with the morphemes {-SHRINK}/{-GROW}. The opening and closing movements of the morphemes {-SHRINK}/{-GROW} are used to represent increase and decrease. Without placement at a particular location, these morphemes are meaningless.

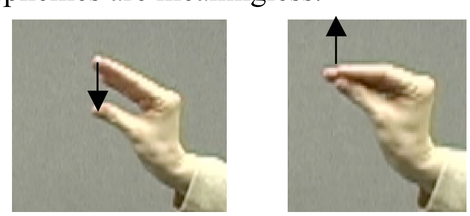


Figure 6: Iconic Morphemes

When placed at a location at the head, the morpheme {-GROW} expresses an increase in knowledge (MIND-GROW). If the morpheme {-GROW} is placed on the chest or close to the heart, the meaning relates to emotional well-being. The morphemes {-

GROW} and {-SHRINK} can also be placed by the ears to symbolize hearing more or less. The diverse use of these iconic morphemes demonstrates the variability of the IS lexicon.

KNOW MIND-GROW

Figure 7: Metaphoric Use of {-GROW}

In the area of interrogatives, on the other hand, IS seems to be still very limited. Out of 99 interrogatives appearing in the present data, 97 are the sign glossed as WHAT. The remaining two are loan signs from ASL (HOW and WHY), are recognized as slips by the interpreter, and corrected with the additional use of the sign WHAT.

Figure 8: IS Interrogative WHAT

This strongly suggests that IS only utilizes one interrogative. It is important to note, however, that the English source text includes a variety of different interrogatives. The examples below shows the English source text and a gloss of the IS translation.

(1) "How do you get financial resources?"
MONEY GIVE GIVE-MONEY POINT WHAT

(2) "What are the sources available to you?"
PRO-1 THINK WHAT WORK SELF WHAT

(3) "[…]who were stay-at-home moms […]"
PRO-1 WHAT

Overall, the IS lexicon shows a variety of highly iconic signs that are very common among all natural SLs, as well as loans specific to single SLs or SL groups. Only a very small number of signs is unique to IS. Generally, the interpreters seem to use the vocabulary in a flexible way, inventing nonce signs where necessary. The use of the one interrogative, on the other hand, suggests that the IS lexicon is limited in its expressiveness by the pragmatic constraints of broad accessibility.

6 IS Structure

This present analysis of structures found in interpreted IS confirms observations made in previous studies and reveals many new aspects. The assumption that IS makes use of many complex structures resembling those of natural SLs is confirmed on all linguistic levels.

Observations on the phonology of IS revealed structures that are as complex as those of natural SLs. The handshape inventory is varied and includes complex handshapes. Phonotactic constrains match rules observed in natural SLs. These observations confirm the assumption by Moody (1989), Locker McKee and Napier (2002) and Webb and Supalla (1994) that the most basic make-up of IS is similar to that of natural SLs. The fact that complex handshapes are rarely used points to the pervasive attempt of the IS interpreter to simplify structures as much as possible.

The number system in IS shows the evolution from an entire iconic system to a more conventionalized one. Smaller numbers are often represented in an entirely iconic way with a one-to-one correspondence between the number of fingers and the number of items that is referenced. This is always true for the numbers between 1 and 10 and in some instances for numbers up to 20. For higher numbers, economic considerations control the structure. Digits are represented in the order of written Arabic numerals. This corresponds with observations for young SLs, specifically Katseff's (2004) work on the number system in Nicaraguan Sign Language (LSN). She observed that with increasing conventionality of the SL in general, number signs become more and more abstract, moving from a one-to-one representation to a representation of digits, and, in some cases, to a more complex and abstract structure (such as the numbers '6' to '10' in ASL). This last step of development was not observed in this IS data and is unlikely to occur, since IS prefers a high degree of iconicity.

The morphological system observed in this IS data is highly complex and in many ways comparable to that of natural SLs. As observed by Moody (1989) and Locker McKee and Napier (2002), IS makes use of space for grammatical purposes extensively. In the present study, the pronoun system and verb system were investigated comprehensively. IS has different forms for personal, possessive and reflexive pronouns. All pronouns use space iconically to indicate referents by directing the sign toward locations or people. Interestingly, the sign PRO-1 ("I"/ "me") seems to be used as the generic person in IS.

IS indicating verbs also use space for grammatical purposes. By directing signs toward an entity and moving toward another, the subject and object of a sign can be identified though spatial references. This phenomenon is described estensively for natural SLs (see Liddell 1980, Engberg-Pedersen 1984). Taub (2001) describes both the pronoun system and indicating verbs as very iconic structures that are common in many SLs. The fact, however, that both the pronoun system and the indicating verbs in IS seem to adhere to certain restrictions (handshapes for the pronouns are lexicalized; only specific signs allow for the spatial reference of trajector and landmark, etc.) points to the complexity of the grammatical properties and a certain level of abstraction within the system.

IS has been researched most extensively on the syntactic level. Webb and Supalla's 1994 study on negation described grammatical facial expressions not only for negation, but also for topicalizations and rhetorical questions. Locker McKee and Napier (2002) observed the use of rhetorical questions and topic markers to divide the source message into smaller units that they hypothesized are easier to process. The present data confirms this function of both topicalization and rhetorical questions. These two grammatical markers are the most frequent in the data. Less frequent, but also occurring, are Wh-questions, relative clauses, yes/no questions and conditionals. The forms of facial expressions seem to correspond largely with the ones described by Liddell (1980) for ASL. Affirmatives and negations are frequently used to modify the meaning of manual signs. No lexical items are necessary to convey consent or disagreement with a proposition. The complexity of the grammatical facial expressions and the wide array of functions they fulfill in IS point to an underlying structure that is maixamally accessible linking facial expressions to message content. Whether the forms of the facial expressions are indeed universal, or rather an influence from specific natural SLs, will have to be researched further.

On a discourse-pragmatic level, many structures were found that correspond to natural SLs structures. One of the most frequently described characteristics of IS that stands out to any observer is the extensive use of role play (see Moody 2002, Locker McKee and Napier 2002, Woll 1990, Rosenstock 2004). In several cases, actions are performed by assuming a role, rather than by desciption. This adds an element of personalization to the discourse that makes it easier for the members of the audience to envision the events. At other times, role play is used to re-enact dialogues or events or to demonstrate the relationship between two or more people.

The techniques used for role play in IS are identical to those documented in natural SLs (see Liddell 2003).

Two structures that have not been described before in the IS literature were frequently employed by the IS interpreters in this study; these are tokens. Tokens are created when signers associate specific locations in space with different concepts. Tokens give interpreters a way to refer back to concepts without having to reiterate explanations. In some sequences, up to ten different tokens are introduced and re-accessed. Comparisons, contrasts, dichotomies of power and other relations can be expressed by specific placements of tokens based on iconic and metaphoric principles. While Moody (1989) and Locker McKee and Napier (2002) have described IS as placing persons in space, the complexity and multilayered nature of the tokens observed in this data suggest that IS interpreters assume that tokens are universally known and used.

On a discourse-pragmatic level, IS uses many techniques to convey information despite the lack of a conventionalized grammar and rich lexicon. The lack of signs for many concepts forces the IS interpreters to expand upon concepts in lengthy explanations. While this is useful in terms of the increased iconic value, these explanations are economically costly because they take a long time. Therefore, in many cases the interpreters limit the explanation of a concept to whatever is immediately necessary. This is illustrated with the introduction of the concept 'loan' in one context as a request for money and then, in another context, as the entire process of receiving money, working and paying it back. The lack of vocabulary also leads IS interpreters to repeat entire discourse sections. Often, the topic of a section is introduced in the beginning and repeated again at the end of an elaborate explanation. This technique might be used to compensate for the lack of syntactic connectors available to the interpreters. In many cases, the interpreter also has to omit information from the source text. A comparison of the ASL source text, English interpretation and IS output would give further insights into the choices the interpreter has to make in terms of expansions and omissions.

7 Conclusion

The analysis presented here shows an extremely complex grammatical system with a rather limited lexicon. As Haiman (1985:535) observes for taboo languages, systems with a limited vocabulary tend to have a more 'cumbersome' grammar and a higher degree of iconic motivation. This is true for IS. The temporal constraints of the discourse setting force the interpreters at times to choose time-saving strategies over iconic motivation. At other times, iconic structures are chosen despite the temporal constraints. The IS interpreters are constantly forced to consider the difficulties of an extended lag time or omissions versus ensuring the understanding of the audience by employing maximally iconic techniques. The iconic elements documented in IS resemble strongly those found in natural SLs (described by Taub 2001). The number of IS structures that resemble natural SLs indicate that there is a connection between the IS grammar and natural SL grammar. Webb and Supalla (1994) speculated that the IS interpreters rely entirely on their native SL grammar in producing IS. A strong argument against this position can be made based on the existence of IS structures that differ from the native SL of the interpreter. A comparison of the native SLs of IS interpreters with their production of IS would provide reliable empirical evidence. The increased use of iconic structure (role play, tokens) in comparison to a natural SLs in the present data seems to suggest strongly that Webb and Supalla's position is not accurate.

Moody (1989) claims that IS interpreters use the most iconic structures of their own SL and superimpose them onto their production of IS. His position leaves unclear where other, more conventional parts of the IS grammar originate. If word order, for example, is assumed not to be iconic, Moody offers no explanation of the source of it in IS. Comparisons of the IS produced by natives of unrelated SLs, such as ASL and Japanese SL, will be required to provide more data to test this hypothesis.

Garretson's (1990) proposal of IS as a structure that is entirely iconic with no influence from a more conventionalized grammar can be rejected on the basis of data found in this study. IS shows several structures where convention is apparent. The use of a QUESTIONMARK sign to emphasize the structure of a sentence as a question only occurs at the end of a phrase, never at the beginning. Since there is no obvious iconic reason to use this particle at the end, this seems to be a conventionalized aspect of IS. The use of a similar particle in Kuwaiti SL is restricted to the beginning of a clause. This further supports the conventionality of the position of the QUESTIONMARK particle in IS. Another area where the sole use of universal, iconic structures is disproved is the SVO structure found both in Supalla and Webb's (1995) study and in this data. Goldin-Meadow and Mylander (1991) found the basic order of constituents in home sign systems to be agent-patient-action. If this is assumed to be a reflection of the maximally iconic structure, than the SVO (agent-action-patient) order found in this data suggests at least some degree of conventionality.

Lastly, Allsop et al. (1994) suggest the grammar of IS to be based on both universally common iconic structures and conventionalized rules. The findings in this study suggest that this is indeed the case. IS does have some conventional structures, as demonstrated above. Yet, the heavy reliance on tokens, surrogates, buoys, DVs, IVs indicates that the IS interpreter in the production of IS assumes a shared knowledge with the audience of all these structures. This suggests that these structures are universally available. This study demonstrates that IS relies on these very iconic structures to an extent far beyond what previous researchers have described.

Researchers in the past have attempted to classify IS as either a pidgin, a Creole or a koine. The complex grammatical structure and limited lexicon described for IS in the past and confirmed in this research rules out a classification as a pidgin. Koines, on the other hand, describe systems that are more complex in their grammatical properties and are often based on language dialects with a similar structural make-up. Koines are developed by neighboring communities that are in constant contact with each other. This is not the case for the community of users of IS. Similarly, creoles are more complex in their grammatical properties than pidgins and might be more similar to IS in that respect. Yet they are characterized by sociological processes that involve a generation of native speakers. This does not apply to IS which is a system with no native signers. Research on contact between two SLs (see Lucas and Valli 1992, Quinto-Pozos 2002) will likely help classify IS more adequately.

As this research has demonstrated, IS continues to evolve. Discourse strategies such as the use of tokens and role play show a level of sophistication similar to natural SLs. On the other hand, the restricted IS lexicon points to areas where IS will likely continue to develop. Future studies will have to show whether the current system suffices to fulfill the demands of international communication in the Deaf Community, whether it will be expanded or replaced by a natural SL.

8 Bibliography

Allsop, Lorna; Woll, Bencie; Brauti, John Martin. 1994. International Sign: The creation of an international deaf community and sign language. *Sign Language Research 1994,* ed. by Heleen Bos and Gertrude Schermer, 171-188. Hamburg: Signum Press.

Garretson, Mervin (ed.). 1990. *Eyes, hands, voices: Communication issues among deaf people*. Silver Spring, Maryland: National Association of the Deaf.

Goldin-Meadow, S., & Mylander, C. (1991). Levels of structure in a langauge developed without a language model. *Brain maturation and cognitive development: comparative and cross-cultural perspectives,* ed. by K. R. Gibson and A. C. Peterson, 315-344. Hawthorn, NY: Aldine Press.

Haiman, John (ed.). 1985. *Iconicity in Syntax: Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24-6, 1983*. Amsterdam: John Benjamins.

Katseff, Shira. 2004. *Number System in SLN (Nicaraguan Sign Language)*. GALA Lecture Series. Gallaudet University. 09/2004.

Liddell, Scott. 2003. *Grammar, Gesture and Meaning in ASL*. Cambridge: CUP.

Locker McKee, Rachel; Napier, Jemina. 1999. *International Sign Interpreting: What can we learn from it?* Unpublished manuscript.

Locker McKee, Rachel; Napier, Jemina. 2002. Interpreting into International Sign Pidign: An analysis. *Sign Language & Linguistics 5:1*. 27-54.

Lucas, Ceil; Valli, Clayton. 1992. *Languages Contact Phenomena in the American Deaf Community*. San Diego: Academic Press.

Moody, Bill. 1989. *Communication International en Milieu Sourd. Le Pouvoir de Signes,* ed. by Alexis Karacostas; Couturier, Lysiane, 89-98. Paris: INJS.

Moody, Bill. 2002. International Sign: A Practitioner's Perspective. *Journal of Interpretation*. 1-47.

Quinto-Pozos, David. 2002. *Contact between Mexican Sign Language and American Sign Language in Two Texas Border Areas*. Ann Arbor, Michigan: UMI.

Rosenstock, Rachel. 2004. *International Sign, Investigating Structure and Comprehension*. Dissertation, unpublished.

Supalla, Ted; Webb, Rebecca. 1995. The grammar of international sign: A new look at pidgin languages. *Language, Gesture and Space*, ed. by Karen Emmorey; Judy Reilly, 333-352. Hillsdale: Erlbaum.

Taub, Sarah. 2001. *Language from the Body. Iconicity and Metaphor in ASL.* Washington DC: Gallaudet University Press.

The British Association of the Deaf. 1975. *Gestuno. International sign language of the deaf.* Carlisle: British Deaf Association.

Webb, Rebecca; Supalla, Ted.1994. Negation in International Sign. *Perspectives on Sign Language Structure: Papers from the Fifth International Symposium on Sign Language Research* [Vol. 1], ed. by Inger Ahlgren; Brita Bergman; Mary Brennan, 173-186. Durham, England: ISLA

Woll, Bencie. 1990. International Perspectives on Sign Language Communication. *International Journal of Sign Linguistics* 1:2. 107-120

A Model for Knowledge Transfer via Linguistic Communication

Yoshihisa Shinozawa
⋆Keio University
Hiyoshi, Kohoku-ku
Yokohama 223-8522, Japan

Akito Sakurai
†Keio University
Hiyoshi, Kohoku-ku
Yokohama 223-8522, Japan
CREST, Japan Science and Technology Agency

Abstract

A model for knowledge transfer and sharing via linguistic communication is proposed. Our analysis of the model suggests that the knowledge transfer requires a mechanism for structure learning which is qualitatively different from the one that enables imitation learning.

1 Introduction

Acquisition of linguistic capacity is considered to be a key factor for the expansion of Homo sapiens across the globe between 100,000 and 15,000 years ago. Knowledge transfer was an essential function that was enabled by the linguistic communication capability, although gossiping, expressing emotions, and others were also very important functions to establish and manage stable societies. Its supporting evidence is a fact that some species other than Homo sapiens, to some degree, have methods to convey emotions and are managing societies; but only Homo sapiens has developed and maintained knowledge. Linguistic communication has been a primary tool.

In this paper, we propose a model for knowledge transfer via linguistic communication and discuss what might be the finishing touch to establish knowledge transfer (or linguistic communication) capabilities.

Our analysis of the model suggests that the knowledge transfer via linguistic communication is made possible by ability to induce finite state automata from the descriptions of the knowledge, where crucial function is structure learning, the one requiring the generalization capability which, we suggested (Shinozawa and Sakurai (2004)), is necessary to acquire syntactic rules. The structure learning is a mechanism qualitatively different from the one which is used in imitation learning now often discussed with mirror neurons (Rizzolatti and Craighero (2004)).

Hurford (2004) argues that “mirror neurons cannot give us any new insight into one of the most crucial features of language, namely the meanings of signs.” We are approaching the same problem (if mirror neurons have something to do with language) from a structural point of view, that is, if sharing of meanings of words were provided by mirror neurons, sharing of knowledge, structures constructed from words and their relations, would not be possible.

2 Formulating knowledge transfer and linguistic communication

2.1 Knowledge representation

Knowledge we have in our brain is nothing less than neural activities in the brain. Since the neural activities are modeled by finite state automata, so is the knowledge.

Brain is a complex system, or more than that. If temporal coding is being used, the brain processes information much differently or in a more complex manner than we imagine. There are, though, many kinds of noise in the brain and if the brain is robust to the noise, information processing by the brain should be equivalent to that by at most finite state automata (see below). Even the brain were not robust to noise, if our common knowledge is the result of communication between us, it should be modeled by finite state automata, since otherwise, it could not be conveyed to others and failed to be common. Finite state automata are learnable in polynomial time but in the limit if the learner could ask membership queries to the teacher, but pushdown automata are not learnable in polynomial time.

Although the neural activity is appropriately modeled by finite state automata, their state spaces being huge, it might be better modeled by smaller Turing machines or quantum computers. Admitting it, we still prefer to model knowledge by finite state automata, since the knowledge that we can communicate to others in limited time is consequently limited in size and complexity.

The hypothesis that the neural activity in our brain is modeled by finite state automata if noise is assumed to exist will not be mathematically proved, since the activity is not well formulated in terms of mathematics. But the fact that recurrent neural networks are at most finite state automata if noise exists as Maass (Maass and Orponen (1998)) proved suggests that the above hypothesis is valid under a wide variety of conditions. Note that recurrent neural networks are not a poor representation method but are more than Turing machines when they are allowed to perform noiseless infinite-precision calculation (Siegelman (1999)).

2.2 Knowledge description by language

For knowledge to be transferred, it must be described somehow. Our best and natural way is to use a languge. Language used to describe and transfer knowledge is a symbolic system, since knowledge transfer may be done by "written" characters only, *i.e.*, without extra-linguistic features such as intonations, facial and body expressions.

We know that the language we use in daily lives is more versatile in expressing things and thoughts than what a moderate size of symbolic system can do. Context, custom, intonations, facial and body expressions are secondary but important media to communicate information. But the role played by language for knowledge transfer is equivalent to the one played by "written" characters to transcript knowledge, since knowledge has been transferred by written characters or by texts. Even if the extra-linguistic features were important, since we are sure to be imperfect in recognizing these features, we can again resort to the Maass' result (Maass and Orponen (1998)) and we can safely say that the descriptions are produced by finite state automata.

A note on describing static knowledge by an automaton. In conventional knowledge representation, knowledge is expressed in, say, a list form (e.g. in LISP). This type of knowledge would be transformed into automaton representation by expressing each atom by a state and sequence by state transition, that is, a traversal of data structure would be into a state transition.

Some dynamic knowledge may not be representable without using context-free grammar or more expressive way. We could, though, approximately represent it by using finite state automaton by restricting, say, the depth of embedded clauses.

In real applications, sentences are an important tool to describe our knowledge about the world. Hotaka (Hotaka (1981)) proposed, implemented, and applied in business field the idea that simple sentences are suited to design schema models for databases.

2.3 Linguistic communication for knowledge transfer

Communication is an endeavor to attain mutual understanding or to build common knowledge among people by using symbols. Linguistic communication is therefore communication with language as a primary media.

We will restrict our attention to two-party communications. For the two parties to have common knowledge, one party should learn the other party's knowledge, recognize the difference between it and his own, and reduce the difference by communications that follow. Therefore, to have common knowledge, knowledge transfer is a basic component. Note that if communication is defined to be one-way information transfer, knowledge transfer is the communication itself.

One's knowledge is not completely transferable to another. It is very common not only in real world but also in our model. It is simply because the knowledge itself is not transferable but only its descriptions are.

We require, then, that if we could communicate as long as we need, knowledge would be transferred completely. The concept is a version of well-known "identification in the limit" introduced by Gold (Gold (1967)). A method identifies a language in the limit if the method identifies the language correctly in some finite time during a course of inductive inferences but it might not notice it. In our version, one party identifies the other party's knowledge in some finite time via communications but it might not know when it happened.

We suppose that the finite state automata we adopt in representing knowledge are of Mealy type (Mealy machines). The inputs/outputs are associated with transitions and not with states, and are identical for each transition. The models are considered to be both acceptors and generators. The input/output symbols are characters, following conventions in automata community, although words are more appropriate when we consider linguistic expressions to describe knowledge. When the automaton is used for producing strings or describing the automaton, these characters are used as output characters; when it is used for accepting strings or testing if a string is properly describing the automaton, the characters are used as input characters. There might be multiple outbound transitions from a state which implies multiple output strings.

3 A mathematical model

Under the above conditions, we could build a few types of models. We mention two variants of two types in the following. In this section an "automaton" refers to a finite state automaton.

1. A model to transfer knowledge. One party has an automaton to represent a piece of knowledge and the other infers it. The party with the knowledge is a teacher and the party without it is a learner when described in a learning paradigm. The teacher runs the automaton by alternating possible transitions in the automaton and produces one string for each learning phase. The learner receives the string and checks if the string is accepted by his automaton. If the learner's automaton does not accept it, the learner modifies the automaton. The learner then runs his automaton, produces a string, and presents it to the teacher. The teacher checks if his automaton accepts the string and returns the answer to the learner.

   The query the learner asks the teacher is called a membership query (Pitt (1989)). Therefore this is a finite state automaton induction with membership queries.

2. A model to establish common knowledge. Two parties have their own knowledge. They communicate to infer the other's knowledge and modify his own knowledge to make them common. As described above, this model may be constructed by combining the models for knowledge transfer.

There are two types of models depending on how we think knowledge is represented.

3. Knowledge is in fact represented (or well-approximated) by automata.

4. Knowledge is represented in a much more complex system, but when we are to describe it, we could only use automaton.

In this paper we only consider the combination of 1. and 3. in the above.

4 An analysis of the model

The problem we consider in the paper is described as below:

Problem 1 *There are a teacher agent A_1 with a finite state automaton M_1 and a learner agent A_2 to infer M_1 with a finite state automaton M_2. A_1 runs M_1 and presents its output string to A_2. A_2 receives the string from A_1 and modifies M_2 based on the result of inputting the string to M_2 if necessary. A_2 runs M_2 and presents its output to A_1. A_1 decides if the string is accepted by M_1 and sends the result to A_2. Can A_2 identify M_1 in the limit by constructing M_2 equivalent to M_1 by repeating the above process ?*

The following theorems hold.

Theorem 1 *Suppose that M_1 is deterministic. If A_1 presents examples faithfully, there exists an algorithm that makes M_2 equivalent to M_1 in polynomial time in the sense of identification in the limit.*

The proof is easily derived from Theorem 1 in Parekh et al. (1998) based on Theorem 3 in Angluin (1981). Since the proofs assume that the set of the terminal symbols are known, we are to modify them to incremental version, which is straightforward.

Also we have stochastic version of the theorem. We may hypothesize that the transitions in the finite state automata be probabilistic (see Clark and Thollard (2004)). In the case, we adopt PAC (probably approximately correct) framework and have analogous results with reasonable restrictions on the target finite state automata.

Theorem 2 *Suppose that M_1 is probabilistic. If A_1 is faithful, there exists an algorithm that makes M_2 equivalent to M_1 in PAC framework in the sense of identification in the limit.*

Note that PAC learnability implies polynomial time complexity. The proof is based on Theorem 5 in Clark and Thollard (2004). Again as before we are to pay attention to the terminal symbol set in the proof.

We should note that the algorithms presented in the proofs are described in symbol processing (or programming) terminology and are not easily transformed into neural network learning algorithms. Although in principle the algorithms could be translated into behaviors of some recurrent neural network, since any computation can be coded into recurrent neural networks with finite precision, the architecture is most probably far away from the neural networks in our brain.

The recurrent neural networks have been tried as a device to learn finite state automaton and achieved some success (Omlin and Giles (2000)). The problem for the recurrent neural networks is that the induced hidden states are only implicitly represented as hidden layer neuron activations and are in general not yet successfully extracted as symbolic rules (Omlin and Giles (2000)). Moreover we do not have a convergence proof. Consequently it is not easy to check the correctness of the learned finite state automaton. In fact, if the target automaton is cyclic, the obtained network tend to miss the cycles.

One more note is that the algorithms used in the proofs are a type of batch algorithm, which recieves all the necessary samples, memorize them, and use them to identify a result of the learning (the result and its candiates are called hypotheses). In contrast to this, an online algorithm receives one sample, modifies its hypothesis if necessary, forget the sample, and cycle the process. Clearly an online algorithm is severely handicapped but is an idealized form of learning machine with fixed amount of memory, which is usually small compared to the examples given; or with fading memory. The above mentioned symbolic algorithms are not online algorithms.

We present here one reinforcement learning algorithm for finite state automata learning, which is an online learning algorithm. An advantage of the algorithm is that it may model knowledge transfer better than symbolic algorithms and a disadvantage is that it is very inefficient.

Let us first explain about the automaton M that we have as a hypothesis. Suppose that we already have a set S of strings *i.e.* descriptions of the finite state automaton to be learned (the target automaton). For simplicity suppose that there is only one final state (accepting state) in our automaton M. Let Σ be the set of the input/output characters, or the terminal symbols. Let $pre(S)$ be the set of the prefix strings of the strings in S. Let the automaton M have the states named $\alpha_i \in pre(S)$, and if $\alpha_i a = \alpha_j$ (the lefthand side is a concatenation of a string α_i and a character a) then M has a transition from α_i to α_j with an input/output character a where α_i and α_j are in S. The empty string λ corresponds the initial state.

Corresponding reinforcement learning RL is defined as follows. The state (α_i, a) of RL is a tuple of the state α_i of M and a terminal symbol $a \in \Sigma$. The action is α_j and the resultant state is uniform-randomly selected from $\{(\alpha_j, a) | a \in \Sigma\}$, which means the environment is stochastic.

The Q-value $Q((\alpha_i, a), \alpha_j)$ is initialized to $-\infty$ (which means the value will not change) when M has a transition from α_i with input/output character a to some $\alpha'_j \neq \alpha_j$. The other Q-values are initialized to 0.

A series of actions starts with the initial state, proceeds with selecting the next action randomly or in ϵ-greedy fashion until reaching the final state. The reward is 0 for all the intermediate transitions, positive if the string produced is correct, and negative if it is wrong.

It is not difficult to see the above reinforcement learning problem RL converges with TD(λ) or Q(λ) algorithm (Dayan (1992)) when $pre(S)$ is a live complete set of the target automaton (Angluin (1981)). The last condition is satisfied if the teacher will not hide a short positive sample (all of which together, expressed informally, constitute a live complete set), *i.e.*, the sample S is large enough.

The next theorem follows:

Theorem 3 *There is a reinforcement learning algorithm that, if A_2 uses, A_2 can infer a deterministic finite state automaton M_2 which is equivalent to the target finite state automaton M_1.*

The algorithm does not give a canonical automaton (*i.e.* an automaton with the minimum number of states among the equivalent ones). To obtain it, we are to run the same algorithms with different S in parallel and choose the smallest among converged ones.

We conclude based on the above arguments that two-party knowledge transfer is possible in the sense of identification in the limit. If we use the above mentioned symbolic algorithm it runs in polynomial time, but may not so if we use an online algorithm.

Note that if we are to transfer knowledge among many parties, we have to cease two-party transfers in halfway. In the case errors may increase along the way of transfers.

5 Comparison with imitation learning

We will discuss briefly some insights obtained from the arguments so far for the possibility of knowledge transfer by linguistic communication.

There have been many discussions on possible relations between language evolution and imitation learning via mirror neurons (Rizzolatti and Craighero

(2004)). There is clear difference, though, between imitation learning of actions and knowledge transfer from our viewpoint.

Although we compare the knowledge transfer to the imitation learning in the following, the results apply to comparison of language acquisition to imitation, since language acquisition is a problem to infer grammars more powerful than finite state grammar, which is a kind of super problem of knowledge transfer.

Firstly, the imitation learning aims at acquiring one best behavior (or, in a more general form, one best policy to achieve one best behavior in possible situations) but knowledge transfer aims at obtaining a best description of a set of infinite, acceptable, but seemingly quite different strings.

Secondly, two models are different in the state descriptions. In the action imitation learning, possible states are described with visual and other sensory measurements (Morimoto and Doya (2001); Schaal (1999)). In knowledge transfer model, no cues of the way to describe the states are given. In both problems, we have to discover or induce intermediate states other than initial and goal states. Those states will be described by sensory imputs in the imitation learning, but by meaningless numbers or id's in the knowledge transfer.

In the action imitation learning, the search spaces for the intermediate states are so huge that it is impossible to search it by brute force and that imitation (together with physical constraints (Harris and Wolpert (1998))) is thought to be a device to overcome the difficulty (Schaal (1999)).

In the knowledge transfer, the search spaces are combinatorially explosive, and cues to overcome the difficulty are only the inputs to the learner.

For imitation learning by robots, reinforcement learning paradigm is often adopted (Morimoto and Doya (2001)). Standard techniques for reinforcement learning try to find out a map from a state description to a best action. Therefore reinforcement learning algorithms are first consideration to solve the imitation learning but are not for the knowledge transfer.

The problem to find out the structure of hidden states is sometimes called hidden state problem or structure learning, which is intrinsically a problem to be solved symbolically, or (as we mentioned above) by enumerating hypotheses one by one from small to larger ones.

In general smaller hypotheses are preferred because of smaller generalization errors and higher probability of guessing correctly (Li and Vitanyi (1993)). For knowledge transfer, this preference is crucial but for action imitation, physical constraints would be suited.

Thirdly, the similarity measures to be defined between candidate solutions are different. Similarity between the orbits is defined naturally, say, by Euclidean distances. Similarity between knowledge is hard to define. It may be defined based on the set difference, which is, though, hard to measure since the sets are basically infinite and the difference of similar strings (elements of the sets) is hard to define to begin with.

In summary, imitation learning of actions via mirror neurons might not be enough to explain language evolution or emergence of linguistic communication, since mechanisms for knowledge transfer, which is a subset of the mechanisms needed for language acquisition, require structure learning which may require generalization, abstraction, or "variblization."

6 Conclusion

We proposed a model for knowledge transfer via linguistic communication. Knowledge representation was modeled by finite state automata and knowledge transfer was modeled by inductive learning of the automata. In the model, the knowledge transfer is possible in the sense of identification in the limit and requires mechanisms different from those for imitation learning via mirror neurons. Therefore, if linguistic communication was used to transfer knowledge that underlay the development of Homo sapiens and our proposed model reflects the essence of the transfer, more sophisticated mechanisms than mirror neurons were necessary for the development.

References

D. Angluin. A note on the number of queries needed to identify regular languages. *Information and Control*, 51:76–87, 1981.

A. Clark and F. Thollard. PAC-learnability of probabilistic deterministic finite state automata. *Journal of Machine Learning Research*, 5:473–497, 2004.

P. Dayan. The convergence of TD(λ) for general λ. *Machine Learning*, 8:341–362, 1992.

E. M. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

C. M. Harris and D. M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394: 780–784, 1998.

R. Hotaka. *Logical Design of Databases (in Japanese)*. Information Processing Society of Japan, 1981.

J. Hurford. Language beyond our grasp: What mirror neurons can, and cannot, do for the evolution of language. In K. Oller and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 297–313. MIT Press, 2004.

Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, 1993.

W. Maass and P. Orponen. On the effect of analog noise in discrete-time analog computations. *Neural Computation*, 10:1071–1095, 1998.

J. Morimoto and K. Doya. Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonomous Systems*, 36:37–51, 2001.

C. Omlin and C. Giles. Symbolic knowledge representation in recurrent neural networks: Insights from theoretical models of computation. In I. Cloete and J. M. Zurada, editors, *Knowledge-Based Neurocomputing*, pages 63–115. The MIT Press, 2000.

R. Parekh, C. Nichitiu, and V. Honavar. A polynomial time incremental algorithm for learning DFA. In *Proceedings of the Fourth International Colloquium on Grammatical Inference (ICGI'98), LNCS 1433*, pages 37–49, 1998.

L. Pitt. Inductive inference, DFAs, and computational complexity. In *Proceedings of the International Workshop on Analogical and Inductive Inference '89, LNCS 397*, pages 18–44, 1989.

G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004.

S. Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3:233–242, 1999.

Y. Shinozawa and A. Sakurai. A role sharing model of language areas. In *First International Workshop on Emergence and Evolution of Linguistic Communication*, 2004.

H. Siegelman. *Neural networks and analog computation: beyond the Turing limit*. Birkhauser Boston, 1999.

Stable communication through dynamic language

Andrew D. M. Smith*

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

Abstract

I use agent-based computational models of inferential language transmission to investigate the relationship between language change and the indeterminacy of meaning. I describe a model of communication and learning based on the inference of meaning through disambiguation across multiple contexts, which is then embedded within an iterated learning model. The dynamic flexibility and uncertainty inherent in the model leads directly to variation between agents, in both their conceptual and lexical structures. Over generations of repeated meaning inference, this variation leads to significant language change. Despite such change, however, the language maintains its utility as a communicative tool within each individual generation.

1 Introduction

All living human languages are constantly changing. Tiny, often barely perceptible, changes in the contexts in which particular words are used, or the way in which they are pronounced, accumulate over generations of use to such an extent that the language itself becomes unrecognisable in only a few generations. The driving force behind historical linguistic change is widely recognised to be linguistic variation (Trask, 1996). In this paper, I explore the relationship between the incessancy of language change and the indeterminacy of meaning, using an agent-based computational model of iterated inferential communication. Inferential communication focuses on the fact that information is not transferred directly between communicants, but rather indirectly: the hearer infers the meaning of a signal from pragmatic insights and the context in which the signal is heard. The uncertainty inherent in this process means that individuals do not necessarily infer the same meanings, leading to differences in their internal linguistic representations. Over generations of inferential communication, small variations may result in significant levels of language change.

Inferential models of language have already been used successfully to model learning conceptual structures and language in tandem (Smith, 2003b) and the effects of psychologically plausible constraints on lexical acquisition (Smith, 2005), but have not yet been used in the detailed study of process of language change. In the experiments presented here, I build on the basic inferential model, embedding it within a successful model of repeated cultural transmission with generational turnover (Smith et al., 2003), in order to explore the nature and extent of language variation and change across many generations of language users.

The remainder of this paper is divided into five parts. In section 2, I explore the twin theoretical foundations on which the model used to perform the experiments is based, namely cultural transmission and the inference of meaning. In section 3, I describe the model in detail, including how agents create meanings, communicate with each other, and infer meaning from multiple contexts. In section 4, I discuss the different kinds of variation which are present in the model, and how these can be measured. In section 5, I present the results of the experiments themselves, and demonstrate that conceptual and lexical variations result in remarkably rapid and significant change to the language itself. At the same time, the language's utility as a successful shared communication system is reconfirmed within each generation. Finally, in section 6, I provide a summary of the paper's main conclusions.

2 Foundations

2.1 Cultural Transmission

Although we are all genetically endowed with the cognitive capacity to learn and use language, the particular languages we actually learn are not stored in our genes, but are instead those which we hear spoken by people in the communities in which we live. Languages are therefore passed on culturally. Recent research into language evolution has focused on this cultural nature of transmission, building models which represent the external and internal manifestations of language as distinct phases in the language's life cycle: individuals produce their *external* linguistic behaviour based on their internal linguistic representations, and in turn induce their own *internal* linguistic representations, or grammars, in response to the linguistic behaviour, or primary linguistic data, which they encounter. Such models of linguistic evolution are known as expression/induction (E/I) models (Hurford, 2002) or iterated learning models (Smith et al., 2003). The cultural nature of these models is captured in the fact that the linguistic input used by one individual to construct its grammar is itself the linguistic output of other individuals. Differences which occur between the internal grammars of individual members of the population occur as a result of the dynamic cultural evolution of the language itself.

Iterated learning models have been used successfully to demonstrate the cultural emergence of a number of structural characteristics of language, notably compositionality (Brighton, 2002) and recursion (Kirby, 2002). These properties arise through repeated cultural transmission when agents must learn a language made up of signal-meaning mappings from a restricted set of data, through a transmission *bottleneck*. Under such conditions, holistic, idiosyncratic rules of grammar can only be successfully transmitted if the specific signal-meaning pair is encountered. Compositional rules, on the other hand, are preferentially produced due to their generalisability, and thus are much more likely to pass through the bottleneck into the next generation (Smith et al., 2003).

2.2 Meaning Inference

It is important to note, however, that many such models of cultural evolution are characterised by the explicit coupling of pairs of signals and predefined meanings. This coupling necessarily leads to the development of syntactic structure which is identical to the predefined semantic structure, and which undermines, to a significant extent, the claims for emergence. In these models, a linguistic utterance consists of the explicit conjunction of a signal and a meaning, and communication involves the direct transfer of this utterance between agents. In communication, then, both the signal and the meaning are simultaneously transferred. As I have shown previously, however, if the meanings are directly transferred, then there is no role for the signals to play, leading to the *paradox of signal redundancy* (Smith, 2003b, 2005): what is the motivation for language users to spend time and energy in learning a symbolic system of signals which provides them no information that they do not already have from the directly transferred meanings?

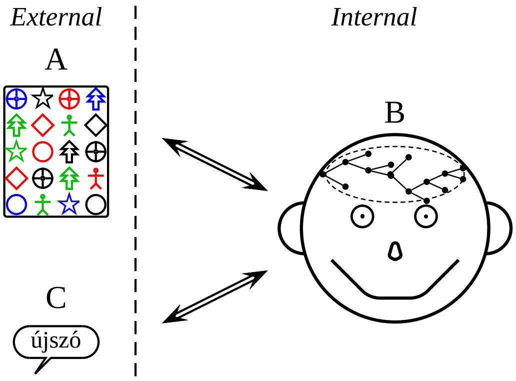


Figure 1: A model of communication which avoids the signal redundancy paradox. The model has three levels of representation: an external environment (A); an internal semantic representation (B); and a public set of signals (C). The mappings between A and B and between B and C, represented by the arrows, fall into the internal, private domain, whose boundary is shown by the dotted line.

The inferential model presented here is motivated to a large extent by avoiding the signal redundancy paradox. This is easily done by recognising that meanings are not directly transferable. Instead, a meaning is encoded into a signal by the speaker, and decoded back by the hearer. Of course, decoupling meanings and signals means that there is now no easy way for agents to associate them with each other, and so we must assume that meanings are *inferred* from some external source. The mere existence of an external world is not sufficient to avoid the signal redundancy paradox, however; we must also insist on a strong demarcation in the model between the external world and the agents' internal representations, as shown in figure 1. The external, or public, domain contains objects and situations which can be potentially accessed and manipulated by all agents, while the internal, or private, domains are accessible only by a particular agent, and contain representations and mappings created and developed by the agent itself.

Signals and their referents are linked only indirectly, mediated via separate associative mappings between themselves and each agent's internal meaning representations. The associative mappings themselves, however, are created individually by each agent through analysis of the co-occurrence of signals and referents over multiple situations, as described in section 3.2 below.

3 The Inferential Model

The E/I models of cultural transmission described in this article, therefore, contain neither a predefined, structured meaning system, nor an explicit link between signals and meanings. Instead, I describe experiments with simulated agents who initially have neither conceptual nor lexical structures, but have the ability to *create* conceptual representations and to *infer* meaning from their experiences. The model contains an external world with a number of objects, which can be objectively described in terms of the values of their abstract features, real numbers generated within the range [0:1]. Agents are provided with dedicated sensory channels, which they can use to sense whether a particular feature value falls within two bounds, and use these to create meanings which allow them to distinguish objects from each other. Agents can also create words to express these meanings and to communicate about the objects. This model is based on that described initially by Steels (1996), in which two agents (a speaker and a hearer) play a series of language games, but is extended in a number of ways. In the following sections, I describe how agents create meanings in response to their interactions with the external world, how they create and use signals to communicate to each other about situations in the world, and how they infer the meanings of signals they receive. Finally, I explain how the inferential model is placed within an iterated learning paradigm, to allow experiments exploring the nature and extent of language change across generations.

3.1 Meaning Creation

Meaning creation occurs as agents explore their environment and try to discriminate objects from each other. In such an exploratory episode, an agent investigates a random subset of objects, called the *context*, with the aim of distinguishing one particular, randomly-chosen *target* object within the context from all the other objects therein. The agent searches its sensory channels for a *distinctive category*, an internal semantic representation which accurately describes the target, but does not accurately describe the other objects in the context. If no such category exists, and so the episode fails, the agent expands its semantic capacity, by splitting the sensitivity range of an existing category into two halves, thereby creating two new categories. Repeated meaning creation in this way results in the development of hierarchical, tree-like conceptual structures where the nodes on the tree represent semantic categories. Nodes nearer the tree root represent more general meanings, with wider sensitivity ranges which cover a greater proportion of the semantic space, while those nearer the leaves represent more specific meanings. Importantly, the simulations contain no pre-specification of which categories should be created, and meaning creation is carried out by each agent individually according to its own experiences. This means that individual agents create different, but typically equally valid, conceptual representations of their world.

3.2 Inferential Communication

Communication follows from a successful discrimination episode. Having found a distinctive category, the speaker chooses a suitable signal from its lexicon to represent it; if none is appropriate, then the speaker creates a new signal as a random string of letters. The signal is then transmitted to the hearer, who also observes the original context from which the speaker derived its distinctive category. Importantly, however, neither the distinctive category nor the target object to which it refers are ever identified to the hearer.

Hurford (1989) developed dynamic communication matrices of transmission and reception behaviour to model the evolution of communication strategies, and showed that bidirectional, Saussurean mappings between signals and meanings are essential in the development of viable communication systems. Oliphant and Batali (1997) extended this model to show that the best way to ensure continuing increases in communicative accuracy is for speakers to always choose signals based on how they are interpreted by the rest of the population. Their algorithm, however, requires agents to be able to have direct access to the internal representations of other agents. In order to avoid this mind-reading, I have used a modified version of the algorithm, *introspective obverter* (Smith, 2003b), in which the speaker chooses the signal which it would be most likely to interpret correctly, given the current context. Because the speaker cannot access the interpretative behaviour of the other agents, signal choice is based on the speaker's own interpretative behaviour.

Once provided with a signal, but without any information about the meaning or the object to which it refers, the hearer must infer the meaning from the information in the context, and from its previous experience of the signal in other contexts. Inference takes place through *cross-situational statistical learning* (Smith and Vogt, 2004). In every situation in which a word is encountered, the hearer creates a list of *semantic hypotheses*, or every possible meaning which could serve as a distinctive category for *any* single object in the current context. Each of these meanings is then associated with the signal in the hearer's internal lexicon. The lexicon contains a count of the co-occurrence of each signal-meaning pair $< s, m >$, which is used to calculate the conditional probability that, given s, m is associated with s. The hearer simply chooses the meaning with the highest conditional probability for the signal it receives and assumes that this was the intended meaning.

If the hearer's chosen meaning identifies the same object as the speaker's initial target object, then the communicative episode is deemed successful. Communicative success is therefore based on *referent identity*: there is no requirement for the agents to use (or even to have) the same internal meaning, but they must identify the same external referent. Furthermore, neither agent receives any feedback about the communicative success of the episode, so the only information available for use in the inferential process is the co-occurrence of signals and referents across multiple contexts. This method of cross-situational inferential learning is similar to the method proposed by Siskind (1996), but differs from it most fundamentally in that the set of possible meanings over which inferences are made in the model presented here is neither fixed nor predefined, but is instead dynamic, and in principle infinite.

Previous experiments using cross-situational statistical learning show that the method is powerful enough for agents to learn large lexicons, and that agents with different conceptual structures can communicate successfully. The time taken to learn a whole lexicon is primarily dependent on the size of the context in which each item is presented (Smith, 2003a; Smith and Vogt, 2004), while communicative success is closely related to the level of inter-agent meaning similarity (Smith, 2003b). However, if agents are endowed with psychologically motivated interpretational biases to aid inference, such as mutual exclusivity (Markman, 1989), then even agents with very dissimilar conceptual structures can communicate successfully (Smith, 2005).

3.3 Iterated Inference

In order to explore how languages change over generations, the inferential model is then extended vertically into a traditional iterated learning model with generational turnover (Smith et al., 2003). It is helpful in this regard to consider the speaker as an adult, and the hearer as a child. Each generation consists of a number of exploratory episodes, in which both agents explore the world individually and create meanings to represent what they find, followed by a number of communicative episodes, in which the adult communicates to the child. At the end of a generation, the adult is removed from the population, the child becomes an adult, and a new child is introduced. The language which was inferred in the previous generation by the child becomes the source of its own linguistic output in the subsequent generation, when it is an adult. This process of generational turnover is then iterated a specified number of times.

4 Variation

It is well recognised that language change is driven by various kinds of variation in language communities (Trask, 1996). In the inferential model I have sketched above, there are two main sources of variation, which I will call *conceptual* and *lexical*. In the following sections, I will describe the source and effects of both types of variation, examples of which can be seen in figure 2. Taken from a representative simulation, this shows an extract from the conceptual and lexical structure of an adult and a child from the same generation. Each agent actually has five sensory channels on which conceptual structures are built, but only one of these channels is shown in figure 2.

4.1 Conceptual Variation

The independent creation of conceptual structure based on individual experience leads inevitably to variation in agents' conceptual representations, both because an agent's response to a certain situation is not deterministic, and because the experiences themselves differ between agents. The relative similarity of two agents' conceptual representations can be quantified by measuring the tree structures built on each sensory channel, then averaging across each sensory channel (Smith, 2003a). If $k(t, u)$ is the number of nodes which two trees t and u have in common,

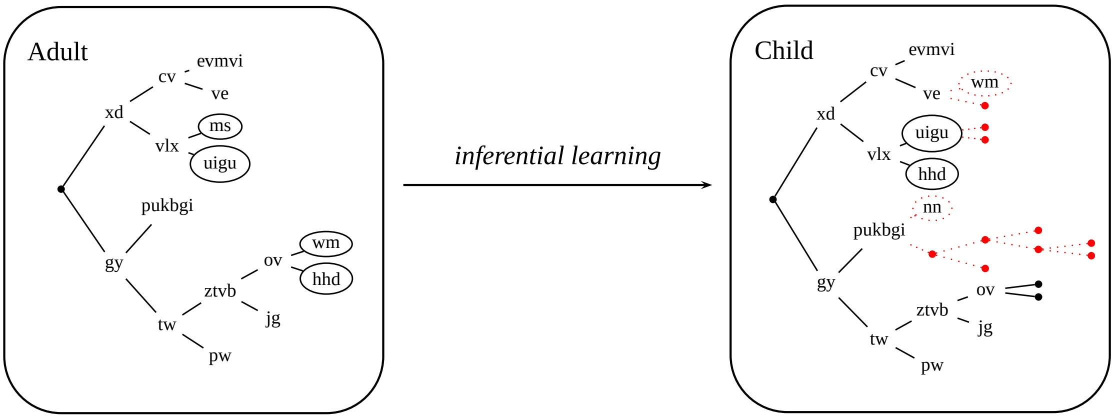


Figure 2: Extract from the internal structures of two agents, showing variation in both conceptual and lexical structures. The conceptual structures are shown by hierarchical tree structures, each node of which represents a different meaning. Conceptual variation, where meanings have no corresponding equivalent in the other agent's conceptual structure, is marked with dotted lines and colour. Lexical structures are represented by the words attached to the nodes, which signify the agent's preferred word for the meaning; empty nodes have no preferred word. Lexical variations, where the agents disagree on the meaning of a word, are circled.

and $n(t)$ is the total number of nodes on tree t, then the similarity $\tau(t, u)$ between trees t and u is:

$$\tau(t, u) = \frac{2k(t, u)}{n(t) + n(u)}.$$

By averaging this measure across all sensory channels, we can produce an agent-level measure of overall conceptual similarity. If a_{ij} identifies the tree on channel j for agent i, and each agent has c sensory channels on which they develop conceptual structure, then the meaning similarity $\sigma(a_1, a_2)$ between agents a_1 and a_2 is:

$$\sigma(a_1, a_2) = \frac{1}{c}\left(\sum_{j=0}^{c-1} \tau(a_{1j}, a_{2j})\right).$$

In figure 2 above, we consider only the nodes on the trees themselves, without reference to the words attached to them. Nodes which have no equivalent in the other agent's conceptual structure are marked with dotted lines and colour. This shows clearly that, although similar, the agents have developed different tree structures: the child has created additional conceptual structure in three different places.

4.2 Lexical Variation

The inherent uncertainty in the process of meaning inference through cross-situational learning also produces variations in the lexical associations made by the agents. Not only are the inferred meanings dependent on the particular conceptual structure the hearer has created, but the associations themselves depend on the particular contexts in which the words are heard. Lexical variation can be measured by considering whether two agents have the same preferred word for any given meaning. An agent's set of preferred words is calculated by sorting a copy of its entire lexicon in descending order of conditional probability (see section 3.2), then mainpulating as follows:

1. find the topmost lexical entry, which is made up of signal s and meaning m.
   (a) store s as the preferred word for m;
   (b) delete all lexical entries containing s.
2. repeat step 1, until the lexicon is empty.

In figure 2, preferred words are represented by the words attached to the appropriate nodes on the tree structure; empty nodes have no preferred word. If adult and child both have the same preferred word for a meaning, then the child has successfully learnt the word, and the lexical item has *persisted* through the generation. Lexical items which do not persist have undergone different kinds of semantic change; these are shown as circled words in figure 2. For example, the words *wm* and *hhd* have not been learnt successfully, despite the relevant nodes in the adult's conceptual structure also existing in the child's structure. In both of these cases, the words are associated with nodes nearer the root of the child's tree than the adult's; because nodes nearer the root of a tree cover a larger area of semantic space, I consider this kind of change as a generalisation. Other kinds of semantic change, such as specialisation and analogy are not discussed further here.

Lexical persistence across the whole of an agent's lexicon is very useful as a broad measure of linguistic change, and can be measured both within and between generations: *intra-generational* lexical persistence is the proportion of the adult's lexicon learnt by the child, while *inter-generational* lexical persistence is the proportion of the original language developed by the adult in the first generation of the simulation which is still intact in the language of the child at the end of the nth generation.

5 Experimental Results

The aim of these experiments was twofold. Firstly, I wanted to verify whether results obtained in previous experiments with an inferential model in a single generation, briefly summarised in section 3.2, would remain valid in a multi-generational model. More importantly, I wanted to measure how languages themselves change over a number of generations, to explore whether languages undergoing rapid language change over successive populations of language users could still be communicatively viable.

5.1 Communicative Success and Meaning Similarity

I have previously shown in mono-generational inferential models that levels of communicative success are closely correlated with levels of meaning similarity between agents (Smith, 2003b). Figure 3 shows results from a typical simulation run over ten generations, each of which is made up of 20,000 episodes. Analyses of communicative success and meaning similarity were calculated every 1000 episodes: communicative success measures the proportion of successful communications over the previous 1000 episodes, while meaning similarity is measured as described in section 4.1.

We can clearly see that levels of meaning similarity and communicative success are again very closely correlated, as expected. In each generation, the communicative success rate rises rapidly at first, as the child successfully learns the meanings of many words through cross-situational inference. The rate then climbs more slowly, as the child tries to deduce the meanings of the remaining words. These represent

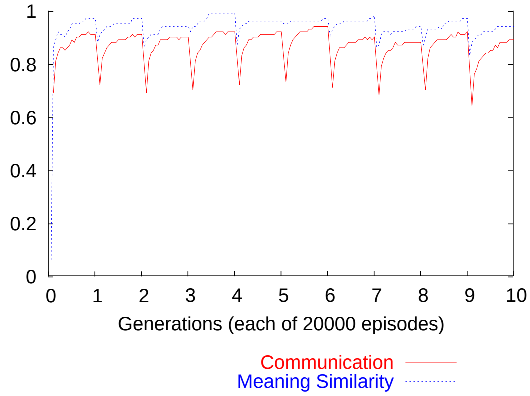


Figure 3: Communicative success and meaning similarity in an iterated inference model. Each generation consists of 20000 episodes.

meanings which are seldom used by the adult as distinctive categories, and so consequently occur relatively infrequently in communicative episodes, which makes the process of disambiguation through exposure in different contexts much slower.

Levels of communicative success and meaning similarity at the end of each generation were also measured, to see if any inter-generational trends were present, but we can clearly see in figure 3 that the levels of communicative success and meaning similarity achieved at the end of each generation were very similar, and in fact no significant inter-generational changes are discernible. This latter results contradicts recent work by Vogt (2003), however, who claims a small increase in inter-generational communicative success in simulations run through his Talking Heads simulator, using a similar model of inferential learning, which he calls selfish games.

5.2 Lexical Persistence

Secondly, I explored changes in the languages themselves over generations of different lengths, measuring lexical persistence to determine the nature and extent of change. Figure 4 shows both inter-generational and intra-generational lexical persistence over simulations of ten generations.

A comparison of the two graphs in figure 4 shows us how the length of a generation (the number of episodes which it contains) affects both measures of lexical persistence. Generations containing 5000 episodes (left) result in intra-generational lexical persistence rates at the end of each generation of between 60 and 70%, but if the generation length is increased to 20,000 episodes (right), then the lexical persistence

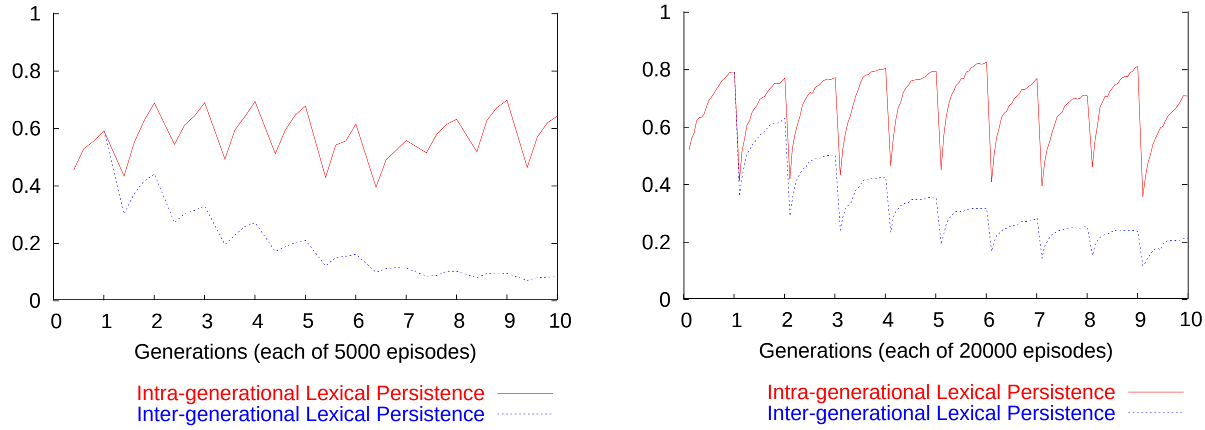


Figure 4: Inter-generational and intra-generation lexical persistence. Each generation consists of 5000 episodes (left) and 20,000 episodes (right).

rates are closer to 80%. Unsurprisingly, given more exposure to the language, the child is able to learn a higher proportion of it successfully. Note, however, that variation in the conceptual structures of the agents provides an effective ceiling for the level of intra-generational lexical persistence, as it is impossible for the child to learn the meaning of a word if the corresponding conceptual structure does not exist in its repertoire.

Figure 4 also shows that there are no significant differences between the levels of lexical persistence obtained within different generations. It is clear, on the other hand, that the rate of *inter-generational* lexical persistence shows a considerable cumulative decline after only a few generations. There are two separate pressures on the language which enforce its relentless erosion over successive generations of cultural transmission through inference, which can be regarded as twin bottlenecks on the language's transmission.

1. Conceptual variation restricts the number of words which can *potentially* persist into the next generation: only words which refer to meanings which are shared are available to be learnt.

2. Lexical variation, or imperfections in inferential learning, further restricts the number of words which *actually* persist into the next generation.

The pressures from these two bottlenecks naturally result in a cumulative decline in inter-generational lexical persistence. These pressures are compounded in subsequent generations, so that even after only a few generations are passed, very little of the original language remains, and we find a language which is changing very rapidly on an inter-generational timescale. Importantly, however, we can see from figure 3 that this rapid language change does *not* affect levels of communication within a single generation, which remain very high.

5.3 Generally Stable

If we investigate in more detail the languages which are used by the child at the end of each generation, we find that there is a distinct pattern to the language change which occurs. Words referring to more specific meanings tend to disappear first, and only more general words tend to survive across multiple generations. There are two obvious reasons for this, both artefacts of the design of the model. Firstly, the Steelsian method of hierarchical conceptual construction forces some order on the meanings which are created: there is no way, for instance, to create a meaning in the depths of a tree without first creating the relevant meanings further up the hierarchical structure. This restriction necessarily means that the more general meanings nearer the root of the tree, are more likely to be *shared* by the agents, and therefore less likely to be excluded from being learnt by the conceptual variation bottleneck. Secondly, agents use a model of communication which follows the maxim of quantity in Grice (1975)'s philosophical model of conversation, by choosing as distinctive categories meanings which provide sufficient information to identify the target object, but are not unnecessarily specific. This means in turn that more general meanings are more likely to be both used by the adult and also *inferred* by the child, and so are much more likely to pass through the second bottleneck on learning.

6 Conclusions

Although the cultural nature of language transmission is becoming more widely recognised, its inferential character is less widely acknowledged. Inferential communication not only provides an explanation for the existence of otherwise redundant signals, but also allows the construction of realistic models of dynamic language, in which uncertainty, variation and imperfect learning play crucial roles.

In this article, I have briefly presented a model of language as a culturally transmitted system of communication, based on the creation and inference of meaning from experience. Individual meaning creation, and the uncertainty inherent in meaning inference lead to different degrees of variation in both conceptual and lexical structure. Conceptual variation and imperfect learning create twin bottlenecks on transmission, which result in rapid language change across generations. Despite this rapid language change, however, within each single generation the language itself remains sufficiently stable to establish and maintain its utility as a successful communication system.

Acknowledgements

Andrew Smith was supported by ESRC Research Grant RES-000-22-0266, and by ESRC Postdoctoral Research Fellowship PTA-026-27-0094. He would like to thank Jim Hurford, Kenny Smith and Paul Vogt for their valuable input and comments.

References

H. Brighton. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54, 2002.

H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics*, volume 3, pages 41–58. Academic Press, New York, 1975.

J. R. Hurford. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77:187–222, 1989.

J. R. Hurford. Expression/induction models of language evolution: dimensions and issues. In E. Briscoe, editor, *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, pages 301–344. Cambridge University Press, Cambridge, 2002.

S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, pages 173–203. Cambridge University Press, Cambridge, 2002.

E. M. Markman. *Categorization and naming in children: problems of induction*. Learning, Development and Conceptual Change. MIT Press, Cambridge. MA, 1989.

M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1), 1997.

J. M. Siskind. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91, 1996.

A. D. M. Smith. *Evolving Communication through the Inference of Meaning*. PhD thesis, Philosophy, Psychology and Language Sciences, University of Edinburgh, 2003a.

A. D. M. Smith. Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, 9(2):175–190, 2003b.

A. D. M. Smith. Mutual exclusivity: Communicative success despite conceptual divergence. In M. Tallerman, editor, *Language Origins: perspectives on evolution*, pages 372–388. Oxford University Press, Oxford, 2005.

A. D. M. Smith and P. Vogt. Lexicon acquisition in an uncertain world. Paper given at the 5^{th} International Conference on the Evolution of Language, Leipzig, 2004.

K. Smith, H. Brighton, and S. Kirby. Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems*, 6(4):537–558, 2003.

L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multi-agent Systems*, Cambridge, MA, 1996. MIT Press.

R. L. Trask. *Historical Linguistics*. Arnold, London, 1996.

P. Vogt. Grounded lexicon formation without explicit reference transfer. In W. Banzhaf, T. Christaller, J. Ziegler, P. Dittrich, and J. T. Kim, editors, *Advances in Artificial Life: Proceedings of the 7^{th} European Conference on Artificial Life*, pages 545–552. Springer-Verlag, Heidelberg, 2003.

What triggers the emergence of grammar?

Luc Steels

*University of Brussels (VUB)
Pleinlaan 2 1050 Brussels
Sony Computer Science Laboratory Paris
steels@arti.vub.ac.be

Abstract

The paper proposes that grammar emerges in order to reduce the computational complexity of semantic interpretation and discusses some details of simulations based on Fluid Construction Grammars.

1 Introduction

There has been a flurry of recent theoretical models trying to explain how and why human languages may have evolved grammatical structures (Hashimoto and Ikegami (1996), Nowak and Krakauer (1999)), and there has been a growing series of computer simulations and robotic experiments applying such models to evolve grounded communication systems in artificial agents (Cangelosi and Parisi (2001), Briscoe(2002), Steels(2003)). The problem of the origins of grammar is obviously a key question in research on the origins and evolution of languages, and it is only when we have clear theoretical models that we can hope to reconstruct the ontogenetic and phylogenetic pathways towards grammatical language.

Most research reported so far views the problem of verbal communication as a coding problem where a meaning M is coded by the speaker into an utterance u and then u is decoded by the hearer to reconstruct the same meaning M. Both processes are a function of the lexicon and grammar, further called the language inventory, of the speaker I_s and hearer I_h, so that $code(M, I_s) = u$ and $decode(u, I_h) = M$.

It is common to argue that syntax arises to make both the size of the inventory and the length of the utterance for a given meaning more optimal (Nowak and Krakauer,1999). Language inventory can be minimised by using a compositional as opposed to a holistic coding. Utterance length can be minimised by coding certain aspects of meaning using syntactic means such as word order or hierarchy. A smaller inventory makes it easier to learn the language and it has therefore been argued that the learning bottleneck (i.e. the fact that language learners are only exposed to a limited number of sentences) encourages agents to choose a compositional as opposed to holistic coding (Smith, et.al., 2003). Although it cannot be denied that syntax has this kind of optimising effect, this paper proposes a different explanation for the role and therefore the emergence of grammar. Specifically, I will argue that the first primary function of grammar (but not the only one) is to optimise semantic interpretation. I will also argue that 'true' grammar only arises when there is an intermediary layer of linguistic categories and constructions as opposed to syntactic structure only.

The paper first defines formally the problem of semantic interpretation and characterises its computational complexity. It then reflects on the nature of grammar and argues that grammar only arises when there is an extra intermediary layer of syntactic and semantic categories that mediates between form and meaning. The paper then explores a peer-to-peer negotiation approach to the origins of grammar, in which grammatical categories and constraints on the use of these categories are progressively built and coordinated by the agents, triggered by the need to optimise semantic interpretation.

2 Semantic Interpretation

Assume a set of *agents* $\mathcal{A}$. Each agent $a \in \mathcal{A}$ is defined as a pair $a =< W_a, I_a >$ where W_a is the agent's world model consisting of a set of facts $W_a = \{f_1, ..., f_n\}$ and I_a is the agent's language inventory, whose structure is defined later. Agents take turns being speaker and hearer and we will assume that they use the same interventory both for coding and decoding. Coding and decoding is in the service of a more encompassing process: producing and understanding. Language speakers are not just uttering sentences without any purpose. They do it because they want to achieve an effect in the hearer. Although there are many possible effects, reflected in the type of speech act implied by the sentence, we will here focus on just one very common communicative goal: The speaker draws attention of the hearer to an object or event in the shared world situation.

After having chosen a topic T_s, the speaker must first conceptualise what meaning M he is going to use to draw attention to T_s. Conceptualisation is a complex cognitive process and appears to be to some degree language-dependent (Talmy, 2000). Here I will just assume that the topic is one of the objects in the speaker's world model W_s and that conceptualisation selects a subset of the facts in W_s: $conceptualise(T_s, W_s) = M_s \subset W_s$. M_s should be such that it uniquely circumscribes the intended topic, which will be the case if the constellation of predicates used in M_s is true for the topic but not for any other object in the world model. Given M_s, the speaker then uses the coding function to produce the utterance: $code(M_s, I_s) = u$

The hearer now uses his own inventory to decode the meaning of the utterance: $decode(u, I_h) = M_h$. Usually it is assumed that $M_s = M_h$, however that is too simplistic. What the hearer obtains from decoding u is an expression with the same predicates as M_s but with variables instead of objects for the arguments (assuming the simplifying case where $I_s = I_h$). The hearer next needs to perform semantic interpretation, which is the process whereby the variables in M_h are assigned values by matching M_h against the world model W_h. The topic intended by the speaker can then be retrieved, thus completing semantic interpretation: $interpret(M_h, W_h) = T_h$.

A simple example will make the need for this extra step clearer. Consider the noun phrase "the red ball" which refers (draws attention) to an object, $o1$. The speaker's conceptualisation has selected two facts about $o1$: $red(o1)$ and $ball(o1)$, and we will write M_s as $[o1|red(o1), ball(o1)]$, to mean 'the object $o1$ such that the two predicates red and $ball$ hold'. When the hearer decodes "the red ball", he obviously does not know yet which object is intended. He is only told that there is something which is red, that this thing is a ball, and that this is what the speaker wants to draw attention to. Formally, M_h is therefore equal to M_s with variables: $[X_1|red(X_1), ball(X_1)]$[1]. The hearer's semantic interpretation process must then match this expression against the hearer's world model and finds that the variable X_1 is bound to $o1$.

Communicative success occurs when the topic identified by the hearer is unique and the same entity in the real world as the topic originally chosen by the speaker:

$$
\begin{aligned}
&conceptualise(T_s, W_s) = M_s \\
&code(M_s, I_s)) = u \\
&decode(u, I_h) = M_h \\
&interpret(M_h, W_h) = T_h \\
&T_s = T_h
\end{aligned}
$$

Although language learners have been argued to receive no or little direct feedback on the nature of the language inventory, they obviously receive plenty of pragmatic feedback on whether the communication was a success or a failure. For example, if you sit at a table and ask for the plate with salmon by saying "the salmon, please", the success of communication is simply reflected by whether you get the salmon or not.

The problem of semantic interpretation is an instance of a so called constraint satisfaction problem (CSP) which has been widely studied in computer science. Each predicate in M_h can be seen as a constraint on its arguments. The domain of possible values is equal to the entities in the world model. A predicate $p_i(X_1, X_n)$ is satisfied for a particular assignment iff the fact obtained by instantiating the variables is part of the given world model. For example, $p_i(X_1, X_2)$ is satisfied for $\{X_1 = o1, X_2 = o1\}$ iff $p_i(o1, o1)$ is an element of W_h. A possible interpretation of M_h is equal to a complete assignment where all variables in M_h are bound in a way that satisfies all the constraints.

The computational complexity of CSP has been thoroughly studied and this allows us to define the computational complexity of interpreting a meaning structure M_h with respect to a world model W_h. Concretely, we are dealing here with a discrete CSP and assume (simplifying) that the number of possible objects in the world model is finite, hence the set of possible assignments of variables d is finite as well. The maximum number of possible assignments for a given meaning M_h with m variables is therefore $O(d^m)$. Searching through this set to find the assignment(s) that are compatible with W_h is exponential in the number of variables.

The following example makes this more concrete. Suppose that the hearer's world model W_h contains the facts:

> ball(o1), ball(o2), hit(o1, o3), hit(o2,o4), box(o3), box(o4), nextto(o3,o6), nextto(o4,o7), green(o6), green(o2), cube(o6), cube(o7), blue(o5), blue(o7)

[1] Variables start with an upper case letter and values with a lower case one.

and that he hears the utterance: "The ball that hit the box next to the green cube". Suppose furthermore that the hearer has a lexicon that maps the content words in this phrase to the corresponding predicates. For example, "ball" adds $ball(X_1)$ to M_h, "hit" adds $hit(X_2, X_3)$ to M_h, etc., so that the phrase is decoded as:

> [X_1 | ball(X_1), hit(X_2,X_3), box(X_5), nextto(X_6,X_7), green(X_8), cube(X_4)]

There are 7 objects in W_h, and 8 variables in M_h, which makes the set of possible assignments equal to $7^8 = 5764801$, a very large number. Many language sentences feature a much larger set of words and involve situations that involve a lot more than 7 objects. So unless a more intelligent method is found for semantic interpretation, communication is not viable.

A first obvious step is to choose an algorithm that does not search by enumerating the set of possible assignments for each of the variables but starts from the predicates in M_h and enumerates only those assignments that actually occur in the world model W_h for each predicate. The computational complexity of semantic interpretation can then be defined in terms of the number of facts in which the same predicate occurs. Let k be the maximum number of facts in the world model that use the same predicate, then the computational complexity of semantic interpretation is $O(k^m)$. This is still exponential in the number of variables, but, assuming a relatively small size of the world model, will be a much smaller number. Concretely, for the example world model given earlier, k is only 2. (There are two boxes, two balls, two hit events, etc.) And so we get $2^8 = 256$ possibilities for M_h. However for realistic world models this is again going to become very large.

3 The role of grammar

The computational complexity of semantic interpretation can be reduced further either (1) by reducing the number of variables in M_h, or (2) by shrinking the set of objects and facts in the world model, which reduces k. Human language users use quite a few devices (linguistic and extra-linguistic) to restrict the context of a conversation and this reduces the domain of the variables and the maximum number of facts that have the same predicate, but I will not elaborate on that aspect here. Instead I focus on the first question, namely how can speakers and hearers reduce the number of variables in the decoded meaning structure? This is precisely where grammar becomes essential.

The key point of this paper is that *the first purpose of grammar is to reduce the number of variables in a decoded meaning structure and hence reduce the computational complexity of its interpretation*. Going back to the example phrase "The ball that hit the box next to the green cube", we see that there is a lot of additional information in this phrase, beyond the lexicon, that communicates equalities between some of the variables:

- "Green cube" forms a noun phrase so that the hearer knows that the predicate green applies to the same object as the predicate cube, $X_8 = X_4$.
- "The ball hit the box ... " is a verb phrase with "the ball" in subject and "the box" in direct object position. This indicates the roles referents play in the hit-event, leading to the conclusion that $X_1 = X_2$, $X_7 = X_8 = X_4$ and $X_6 = X_5 = X3$.

So we have a reduction from 8 to 3 variables and computational complexity of semantic interpretation reduces from *O*(256) to *O*(8). Variables which are constrained to refer to the same object are called equalities.

The issue is not only complexity. Without the additional information that some of the variables introduced by the lexicon have to be assigned to the same values, there would be several semantic interpretations which are all complete. Going back to the example phrase, we see that, there are in fact 2^8 of them (because I constructed the example so that there are two possible assignments for each predicate). However, when taking the additional constraints on variable equalities communicated by syntax into account, only one interpretation remains. So the secondary effect or grammar is also to reduce the number of possible interpretations so that only a unique complete assignment of the variables remains.

4 From Syntax to Grammar

The next question is how natural languages communicate variable equalities. One way is through syntactic structures, based on word order or extra markings. For example, combining the words "red" and "ball" into "red ball" implies that the variables used in $red(X_1)$ and $ball(X_2)$ are equal, $X_1 = X_2$, so that the meaning becomes $[X|red(X), ball(X)]$. Such a patterning could at first be completely ad hoc, which is the case for example in programming languages. To specify the arguments of a procedure or function, programming languages or logic use ordering. For example, the procedure $DrawWindow(W, x, y, z)$, requires 4 arguments to be supplied in a particular order. Note that ad hoc syntactic structures could already have recursive structure, if a group which forms a unit (like “red ball”) is combined into a larger structure (“red ball next to green ball”). In a programming language, there is no further systematicity in syntax. When defining another procedure like MoveWindow, there could be a totally different ordering: $MoveWindow(x2, y2, z2, W)$ or $move-window(W, z2, y2, x2)$, etc., depending on the programmer’s wish. Of course a good programmer will introduce some systematicity in the syntax he is using but the interpreter and compiler know nothing about this.

An experiment in the emergence of syntax in this sense has been carried out by Batali (2002). His syntactic combination rules contain ‘argument maps’ to specify the variable equalities. They are created in an ad hoc fashion as exemplars. Thus, using numbers for the arguments, the individual words usifala, [(snake 1)(sang 1)] and ozoj [(chased 1 2)], are combined into “usifala ozoj” to express (snake 1) (sang 1) (chased 1 2), with the mapping 1:1 for the first word, and 1:1, 2:2, for the second one. Agents negotiate the use of exemplars, based on a lateral inhibition dynamics: Success reinforces the use of certain exemplars and failures discourages their use. Exemplars are reused as much as possible which implicitly creates at least some systematicity but this systematicity is not captured in rules.

Natural languages however impose an additional layer in between the meaning to be conveyed and the final syntactic form. The meaning is reconceptualised in terms of semantic frames such as a TRANSFER-TO-TARGET frame with agent, target and patient and the form is categorised in terms of syntactic categories (like noun, article, etc.), grammatical relations (like subject, determiner), and syntactic patterns (like a Subject-Verb-Direct-Object pattern). The combination of a semantic frame and a syntactic pattern is known as a grammatical construction (see figure 1) (Goldberg,1995). It is only when such a layer of grammatical constructions with syntactic and semantic categories that one can speak about true grammar. It has the obvious advantage of economy and greater expressive power. Constructions in natural language clearly have different degrees of specifity (i.e. idiomaticity), ranging from very idiomatic constructions built around a particular noun or verb, to very general constructions with wide applicability, such as Subject+Predicate+DirectObject (as in ”John gives a book”). Constructions thus form networks where more specific constructions inherit from more general ones and combine with each other to achieve high expressive power. Moreover empirical observations of actual language use shows that the inventory of constructions used by an individual (including adults) is constantly changing. Constructions capture conventionalised patterns of usage, but new patterns develop all the time and others may go out of fashion.

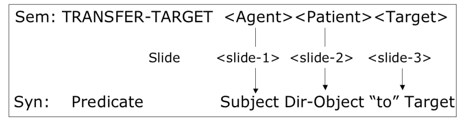


Figure 1: A construction relates a syntactic pattern such as Subject+Predicate+DirectObject+PrepObject with a semantic frame such as TRANSFER-TO-TARGET+Agent+Patient+Target.

To implement all this, we need a formalism which can explicate semantic categorisation rules for reconceptualising meanings into semantic frames, and syntactic categorisation rules that categorise words and syntactic structures. There must also be an explicit representation of grammatical constructions, i.e. associations between semantic frames and syntactic patterns. These constructions should still establish equalities between variables (as Batali’s argument maps) but they will now be more generic and hence applicable to a wide range of situations. If syntactic and semantic categorisations and constructions are explicitly represented in the grammar, then it follows that the agents must have operators for inventing them (as speaker when there are equalities that need to be eliminated), for adopting them (as hearers when there are equalities that the speaker has eliminated) and for aligning them to ensure that the categories and rules of different agents become similar. The next section provides a bit more technical detail on how we have implemented these various aspects.

5 Fluid Construction Grammars

The formalism we have implemented for representing emergent grammars is called Fluid Construction Grammar (FCG) and is related to other computational implementations of construction grammar such as ECG (Bergen and Chang,2003), as well as standard techniques of unification-based grammar employed in computational linguistics today (Pollard and Sag, 1994). Specifically, syntactic and semantic structures are represented as typed feature structures, as shown in figure 2 and 3 for the sentence “Jill slides Jack the block”. Fluidity refers to the goal of being extremely flexible in parsing and production, including when there is no or insufficient grammar or when some rules are violated. In FCG, all rules are bi-directional so that they can be used both for produciton (i.e. constructing an utterance that expresses specific meanings derived through a conceptualisation process from a grounded world model) and for parsing (reconstructing the meaning of an utterance and mapping it back into reality by way of the grounded world model). This is a tough technical requirement which is achieved by viewing grammar rules as constraints and language processing as constraint propagation.

```
unit1
 syn-cat: {sentence,SVOtoO-sentence}
 syn-subunits: {unit2,unit3,unit4,unit6}
unit2
 syn-cat: {person(3d,natural),number(singular,natural),
          subject(unit1,unit2)}
 form: {stem(unit2,"jill")}
unit3
 syn-cat: {predicate(unit1,unit3)}
 form: {stem(unit3,"slide")}
unit4
 syn-cat: {person(3d,natural),number(plural,natural),
          direct-object(unit1,unit3)}
 form: {stem(unit5,"block")}
unit6
 syn-cat: {person(3d,natural),number(singular,natural)
          prep-object(unit1,unit6)}
 form: {stem(unit4,"jack")}
```

Figure 2: Syntactic structure after application of the TRANSFER-TO-TARGET construction. There is a unit for each word and for combinations of words. The syntactic categories as well as the properties of the surface form are represented as predicates over units.

```
unit1
 sem-subunits: {unit2,unit3,unit4,unit6}
unit2
  referent: {obj1}
  meaning {jack(obj1),status(obj1,single-object),
           discourse-role(obj1,external)}
unit3
  referent: {ev1}
  meaning: {slide(ev1,true),slide-1(ev1,obj1),
            slide-2(ev1,obj2),slide-3(ev1,obj3)}
  sem-cat: {transfer-to-target(ev1),agent(ev1,obj1),
            patient(ev1,obj2),target(ev1,obj3)}
unit4
  referent: {obj2}
  meaning: {block(obj2),status(obj2,single-object),
            discourse-role(obj2,external)}
unit6
  referent: {obj3}
  meaning: {jill(obj3),status(obj3,single-object),
            discourse-role(obj3,external)}
```

Figure 3: Semantic structure built up alongside the syntactic structure shown in the previous figure. It contains bits of meaning as well as semantic categorisations necessary for the application of the grammatical construction (in the slot SEM-CAT).

FCG rules contain a left pole and a right pole and are activated and applied through unification. An example of a grammatical construction is shown in figure 4. The left pole constrains the semantic side and the right pole the syntactic side. Other rules will expand the semantic and the syntactic structure with descriptions so that this rule can be applied. For example, there will be a semantic categorisation rule that re-conceptualises a slide-event with its various roles (as in John slides the book to Mary) into a TRANSFER-TO-TARGET event. Producing and parsing are totally analogous, the only thing which changes is the direction of rule application.

```
def-cons transfer-to-target-construction
  ?top-unit
    sem-subunits:
      ?event-unit,?agent-unit,?target-unit,?patient-unit
  ?event-unit
    referent: ?event
    sem-cat: transfer-to-target(?event),agent(?event,?agent)
       patient(?event,?patient),target(?event,?recipient)
   ?agent-unit
     referent: ?agent
   ?patient-unit
     referent: ?patient
   ?target-unit
     referent: ?recipient
  <-->
   ?top-unit
     syn-cat: SVOtoO-sentence
     syn-subunits:
      ?event-unit,?agent-unit,?patient-unit,?target-unit
    ?event-unit
      syn-cat: predicate(?top-unit,?event-unit)
    ?agent-unit
      syn-cat: subject(?top-unit,?agent-unit)
    ?patient-unit
      syn-cat: direct-object(?top-unit,?patient-unit)
    ?target-unit
      syn-cat: prep-object(?top-unit,?target-unit)
```

Figure 4: Example of a construction which relates a TRANSFER-TO-TARGET frame to a Subject+Verb+Direct-Object+to+Prep-Object pattern

Agents in our simulations of grammar emergence create categories and constructions in order to reduce the computational complexity of semantic interpretation and align these categories and constructions based on the outcome of the language game. We summarise the main principles of these simulations and refer to Steels (2005) for more detail.

Suppose the speaker has a target meaning M_s which he wants to communicate to refer to a topic, and he can use his lexicon (and maybe already a partial grammar) to code that meaning into an utterance u. But before sending u to the hearer, the speaker can first determine the complexity of semantic interpretation by *re-entrance*: The speaker decodes u (using his own lexicon and grammar) to yield a meaning M_s', and then tries to interpret M_s' against his own world model W_h. This gives a set of possible bindings and possibly a set of equalities. If there are equalities, the speaker knows that additional grammar should be added. Conversely, if the hearer attempts to interpret his interpretation of an utterance u and obtains a possible referent T_h (possibly after additional interaction with the speaker if there was a failure), then he also has a set of bindings and a set of equalities. If there are equalities, the hearer can interpret the additional syntactic information present in the utterance as a reasonable hypothesis that this information is intended to show how the equalities can be resolved.

We discuss first an example how a specific idiomatic construction is generated. Suppose that the speaker wants to express the following fall event: 'fall(ev1), fall-1(ev1,obj1), ball(obj1)'. Assume that the speaker has already lexical rules for "fall" and "ball", leading to the semantic and syntactic structure in figure 5. No grammar is involved yet.

```
unit1
 sem-subunits: {unit2,unit3}
unit2
  referent: {obj1}
  meaning: {ball(obj1)}
unit3
  referent: {ev1}
  meaning: {fall(ev1),fall(ev1,obj1)}
```

```
unit1
 syn-subunits: {unit2,unit3}
unit2
  form: string(unit2,“ball”)}
unit3
  form: string(unit3,”fall”)}
```

Figure 5: Semantic (left) and syntactic (right) structure after applying lexical rules for "ball fall".

If the speaker re-interprets himself the resulting sentence "ball fall" using his own lexicon, he comes up with the following meaning: 'fall(?ev1), fall-1(?ev1, ?obj1), ball(?obj2)'. If this is matched against the original meaning, the equality ?obj1 = ?obj2 becomes apparent. So if this equality would become expressed grammatically, the communication would become more precise and the risk of failure decreases. The speaker invents a construction for this purpose in two steps.

The first step is to combine the structures derived from the lexicon, introduce variables for all units and entities involved, and add the precedence relation occurring in the sentence, which was arbitrary but now becomes rule-governed. Slots need to contain the specified elements but also could contain other ones. This gives the result shown in figure 6. Note that the variable used with the predicate ball (i.e. ?obj1) is the same as in fall-1. This is the way that the equality will get established when the rule is applied.

This construction does the job in the sense that when "ball fall" is seen, the lexicon contributes the various predicates to the meaning and the construction establishes the right equality. However it is completely ad hoc, so a more general operation should take place, which generalises the meaning and the form by stating the constraints in terms of semantic and syntactic categorisations. The result is shown in figure 7.

The relation between the semantic categorisations and the meaning predicates now needs to be translated into a sem-rules (shown in figure 8). These

```
def-con ball-fall
 ?unit1
   sem-subunits: ?unit2,?unit3
 ?unit2
   referent: ?obj1
   meaning: ball(?obj1)
 ?unit3
   referent: ?ev1
   meaning: fall(?ev1),fall-1(?ev1,?obj1)
<-->
 ?unit1
  syn-subunits: ?unit2,?unit3
  form: precedes(?unit2,?unit3)
 ?unit2
   form: string(?unit2,“ball”)
 ?unit3
   form: string(?unit3,”fall”)
```

Figure 6: The first step in inventing a construction is to perform a kind of lambda-abstraction, introducing variables for units and entities.

```
def-con object-move
 ?unit1
   sem-subunits: ?unit2,?unit3
 ?unit2
   referent: ?obj1
   sem-cat: sem-cat1(?obj1)
 ?unit3
   referent: ?ev1
   sem-cat: sem-cat2(?ev1),sem-cat3(?ev1,?obj1)
<-->
 ?unit1
  syn-subunits: ?unit2,?unit3
  form: precedes(?unit2,?unit3)
 ?unit2
   syn-cat: syn-cat1(?unit2)
 ?unit3
   syn-cat: syn-cat2(?unit3)
```

Figure 7: The second step in inventing a construction is to replace specific meaning and form predicates with semantic and syntactic categorisations.

rules are easily constructed by taking the relevant part of the meaning slot in the original semantic structure and linking it to the sem-cat slot in the construction. These categories are still ad hoc in the sense that they have only one member, but the categories progressively become richer as new elements are declared to be members of them, so that the extent of sem-cat1 becomes something like 'the set of objects which can participate in physical movement events', sem-cat2 becomes 'the set of events that involve such physical movement', and sem-cat3 'the patient involved in this physical movement'.

```
def-sem sem-cat2+3
 ?unit
   referent: ?ev1
   meaning: fall(?ev1),fall-1(?ev1,?obj1)
 -->
 ?unit
   sem-cat: sem-cat2(?ev1),sem-cat3(?ev1,?obj1)
```

Figure 8: A semantic categorisation rule for some of the semantic categories in the construction shown in figure 7

The relation between the syntactic categories and the predicates describing aspects of form is expressed in syn-rules. An example is shown in figure 9. The form constraint is repeated in the left pole to make the rule bi-directional. Again these categories are at this moment ad hoc, having the specific words "ball" or "fall" as only members, but as the construction gets re-used, the category becomes broader and they will become similar to the parts of speech in natural languages.

```
def-morph "ball"
  ?unit
    syn-cat: syn-cat1
    form: string(?unit, "ball")
 <-->
  ?unit
    form: string(?unit, "ball")
```

Figure 9: A syntactic categorisation rule for "fall"

The hearer goes through exactly the same sort of operations for constructing his own grammatical rules. The hearer detects equalities based on the predicates supplied by the lexicon which are matched against the specific situation in the shared environment to yield a set of bindings and possibly equalities. Every rule in FCG has a strength which reflects how much success the rule has had in the past. The strength is updated using lateral inhibition dynamics, already used for the lexicon Steels(1996): Successful application reinforces the rule and failure causes damping. This leads to a gradual self-organised coherence of the agents' repertoires.

Language users should try to optimise their inventories by re-using as much as possible existing constructions to cover new situations. This has two advantages: economy of memory, because fewer rules need to be stored, and optimisation of processing because fewer rules need to be considered. But re-use is also beneficial to speed up learning. If there is already a construction which is more or less doing the job, then the hearer can use that construction as a basis to help guess the meaning and learn more about the syntactic and semantic categories of the speaker. In line with the (embodied) cognitive linguistics approach, we argue that grounding should play a major role in deciding to re-use a construction.

Some cases are relatively straightforward. If there is another fall event but now involving another object, say a block, then "block" can simply be categorised as syn-cat1 and the predicate block as sem-cat1, so that the OBJECT-MOVE construction shown in figure 7 becomes applicable. However other cases are not so straightforward. Suppose that a new event has to be categorised (e.g. 'slide(?ev5), slide-1(?ev5,?obj6), slide-2(?ev5,?obj7)'). The already existing instances of a category (in the example above this is so far only the fall-event with sem-cat2 and sem-cat3) can be compared to the new event by examining the state transition networks that are used for the recognition of each event. The primitive events and event combinations of each event are paired together with the entities that play specific roles in each event. Based on this comparison, a measure of category membership can be computed in a straightforward manner, to find the category whose instances are closest to the new event to be expressed. Thus, besides a 'patient' that is undergoing movement (sem-cat3), an agent is involved in a slide-event. The entity playing this role participates in different primitive events than the patient and hence the corresponding predicate would not fit very well with sem-cat3. There are still other ways in which constructions can be re-used. For example, if there is already a construction like the one shown in figure 7 it could be specialised with additional roles, e.g. to express the agent of the move-event or the manner of movement.

These learning mechanism proposed are 'constructivist' Tomasello and Brooks(1999) in the sense that they are not derived from statistical clustering but imposed by language users and then possibly adopted as consensus. At first the categories are ad hoc and have only a single entity as its member, but as constructions are re-used, more instances are added to the category and so they are getting a richer content. The instance-based learning of categorisation results in the prototype behavior also seen with the linguistic categories found in human natural languages.

6 Conclusions

This paper argued that reducing the computational complexity of semantic interpretation, and hence the chance of communicative success, can be the main driving force for getting a population of agents to develop grammar. It argued also that 'true' grammar only arises when syntactic and semantic categories are used and grammatical constructions to have a more abstract mapping between form and meaning. We do definitely not argue that this is the only use of grammar. In fact when second order predicates become used (i.e. predicates that have other predicates as argument, such as "very" in "very good") there is a second important reason for introducing grammar, namely that the grammar specifies how a predicate needs to be used. Much further work needs to be done to carry the computational simulations forward, and there is no doubt that the operators we have used so far need to be extended with more powerful mechanisms for the invention of new grammar. Chang and colleagues (Chang and Maia,2003) have recently presented computer simulations of such learning processes based on empirical data of child language acquisition and Bayesian learning mechanisms. The perspective adopted here is along similar lines, although we use an abductive learning approach.

Research funded by Sony CSL with additional FWO funding in the ESF-OMLL program, and the EU FET-ECAgents project. The author is indebted to Joachim de Beule, Nicolas Neubauer, and Joris Van Looveren, who have made important contributions, particularly to implementation issues.

References

[1] Batali, J. (2002) The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In (3).

[2] Bergen, B.K. and N.C. Chang. Embodied Construction Grammar in Simulation-Based Language Understanding. Technical Report 02-004, International Computer Science Institute, Berkeley.

[3] Briscoe, T. (ed.) Linguistic Evolution through Language Acquisition: Formal and Computational Models. Cambridge University Press, Cambridge, UK.

[4] Cangelosi, A. and D. Parisi (eds.) (2001) Simulating the Evolution of Language. Springer-Verlag, Berlin.

[5] Chang, N. C. and Maia, T. V. (2001) Grounded Learning of Grammatical Constructions. In 2001 AAAI Spring Symposium on Learning Grounded Representations.

[6] Goldberg, A.E. (1995) Constructions: A construction grammar approach to argument structure. University of Chicago Press, Chicago.

[7] Hashimoto, T. and T. Ikegami (1995) Emergence of netgrammar in communicating agents. Biosystems 38, 1-14.

[8] Nowak, M.A. and D.C. Krakauer (1999) The evolution of language. Proc.Natl.Acad.Scie. USA. Vol 96, pp. 8028-8033, July 1999.

[9] Pollard, C. and I. Sag 1994. *Head-driven phrase structure grammar.* Center for the Study of Language and Information of Stanford University. University of Chicago Press, Chicago.

[10] Smith, K., Kirby, S., and Brighton, H. (2003). Iterated Learning: a framework for the emergence of language. Artificial Life, 9(4):371–386.

[11] Steels, L. (1996) Emergent Adaptive Lexicons. In: Maes, P. (ed.) (1996) *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, The MIT Press, Cambridge Ma.

[12] Steels, L. (1998) The origins of syntax in visually grounded robotic agents. Artificial Intelligence, 103:1–24.

[13] Steels, L. (2003) Evolving grounded communication for robots. Trends in Cognitive Science. Volume 7, Issue 7, July 2003 , pp. 308-312.

[14] Steels, L. (2005) The role of Construction Grammar in Language Grounding. Artificial Intelligence Journal, Vol 31.

[15] Talmy, L. (2000) Toward a Cognitive Semantics: Concept Structuring Systems (Language, Speech, and Communication) The MIT Press, Cambridge Ma.

[16] Tomasello, M. and P.J. Brooks 1999. *Early syntactic development: A Construction Grammar approach* In: Barrett, M. (ed.) (1999) *The Development of Language* Psychology Press, London. pp. 161-190.

‘Needs only analysis’ in linguistic ontogeny and phylogeny

Alison Wray
Centre for Language and Communication Research
Cardiff School of English, Communication & Philosophy
Cardiff University
Humanities Building
Colum Drive, Cardiff, CF10 3BU, UK
wraya@cf.ac.uk

Abstract

Recently, linguists from several quarters have begun to unpack some of the assumptions and claims made in linguistics over the last 40 years, opening up new possibilities for synergies between linguistic theory and the variety of fields that engage with it. A key point of exploration is the relationship between external manifestations of language and the underlying mental model that produces and understands them. To what extent does it remain reasonable to argue that all humans ‘know’ certain things about language, even if they never demonstrate that knowledge? What is the status of knowledge that is only stimulated into expression by particular cultural input? Many have asked whether the human’s linguistic behaviour can be explained with recourse to less innate knowledge than Chomskian models traditionally assume. But to what extent might it be appropriate, in addition, to move away from the quest to model full systematicity at all? This paper proposes that models that generate an untidy and only partially rationalised product may a better match for reality than is often supposed. The implications for simulation work are extensive: a reformulation of target outcomes offers new possibilities for characterising starting states and processes.

1 Introduction

Linguistics in the pre-Chomskian era was in some ways naïve, even prejudiced, but in other ways it was much more open to the exploration of what variation between languages means, and what might underlie such variation. Revisiting the writings of Jakobson, Jespersen, Firth, Bloomfield, Saussure and others can refresh our perceptions of language (e.g. Wray 2002a, chap.1), particularly if we ourselves have been heavily influenced by the Chomskian tradition. We need to move towards a mature synergistic approach that can evaluate the various contributions made to linguistic research over the past century or more, scrutinise points of difference, and identify and examine common assumptions. This is beginning to happen. To some extent language-focussed AI research is fuelling this activity, but beyond that, it has much to gain from engaging with the moving tide of questions about language, many of which can be answered empirically. Amongst the issues that are being examined, not least in response to the difficult questions that must be addressed in language evolution research, are the uniformity of language and of fundamental linguistic knowledge, and what, precisely, is it is that is innate in relation to language.

Approaches to simulation vary, but all need to specify and manipulate a starting state and/or process in order to observe the effect on an end state. At least one of these elements must be defined by and/or evaluated relative to some external reference point, such as a current model of learning, the real world target of the simulation, etc. Pre-specifications of the starting state must be scrutinized, justified and where possible minimized, since they build in features that might otherwise have been explained as a product of the process. It follows that language simulators seek clear answers to questions about the phenomena they model: if, in a model of language evolution, X (Y) $\rightarrow$ Z, then what do linguists, psychologists and others consider reasonable definitions of X (starting state) and Z (end state), such that versions of Y (process) can be explored?

Simulating language makes heavy demands if we construe language as a complex dynamic system with many interacting parts[1]. Simulation research to date demonstrates that even quite simple processes can result in a level of complexity in the end state that it was previously assumed must be subject to specification in the starting state. Such discoveries, however, do not in themselves guarantee that the simulation is

[1] Simon Kirby (personal communication).

a close match for the phenomena it models. Furthermore, linguists have tended to believe that language is fundamentally so complex that it is unlikely to render its secrets easily in simulation studies. Although much is being done to establish how the basic building blocks of language could emerge – sound systems, meaning, basic thematic and grammatical roles, etc. – it remains unclear whether it will ever be possible to write a program that can learn a language, or generate linguistic material, in a way comparable (in outcome and/or process) to that of a human.

With regard to this holy grail, one line of exploration (e.g. Kirby et al 2004) is to establish whether the features of language that are held to be universal need to be pre-specified in a UG, or whether they can emerge on the basis of input. The account developed here is consistent with the latter position, but I also propose that the individual's language knowledge may be more patchy than linguistic theory has generally assumed – if this is the case, it may be easier to model than we thought. I shall explore two main themes. Firstly, some of the complex features of language may be 'universal' only in a secondary sense – bound to manifest in a particular form, but not actually bound to manifest. If so, then although they must still be explained, their place in models of language evolution or acquisition can be marginalized. Secondly, accounting for the presence in language of troublesome elements such as irregularity and semi-regularity may be less a question of how such features are generated than how and why they are tolerated. The theoretical model I shall describe accounts for complexity in a way that brings with it interesting opportunities for simulation studies.

2.0 All languages are equal (but are some more equal than others?)

The strong version of the uniformitarian position holds that since all languages are defined by the design of the human brain, which is the same for all modern humans, all languages must be equal in their complexity (see Newmeyer 2002). Those who take this position set a parameter for simulation studies: if the product is always of the same order, the starting state and process must, between them, account for that fact. As Newmeyer points out, in the absence of contrary evidence, some kind of uniformity in language capacity must be assumed to exist, for "nobody – at least one would hope nobody – has claimed that there exists a language for which subordination is literally *impossible*" (p.369). However, a weaker version of the uniformitarian position is possible.

Newmeyer opens up the possibility that the uniformitarian view can, and should, accommodate certain observable tendencies in language, specifically, directionality in language change, e.g. from verb-final to verb-medial, increasing morphophonemic and phonological complexity, increasing grammaticalisation, reduction in deictic complexity, and reduction of marked structures. He summarises a number of studies, including Nettle's simulation work, that variously suggest influences on linguistic structure from culture, group size, and language contact. He observes that if there is, as considerable evidence suggests, variation in the extent to which certain linguistic features are expressed, this undermines the strong uniformitarian view that all languages are of equal complexity, but does not challenge the weaker version, that all humans possess equal underlying linguistic capabilities.

He concludes that "if grammar is tailored to the needs and properties of language users (to whatever degree), and language users now are not what they used to be, then it follows that grammar is probably not what it used to be" (p.369). It is significant that he specifies "needs and properties of language users" as a determining factor in how the language is shaped. If, as the evidence suggests, one feature that is variable under conditions of user "needs and properties" is grammatically expressed subordination, that will automatically impact on the opportunities for certain supposed realisations of UG to be expressed, such as Subjacency. It follows that while we still have to account for shared underlying linguistic capacities in *Homo sapiens* we may not invoke as evolutionary pressures any features of language that are contingent on socio-cultural or other factors not extant in prehistoric times. Care must be taken, of course. While we can make some reasonable guesses about group size, inter-group contact, ecology and food availability in pre-modern or early modern man, we cannot be sure about the precise circumstances at vital moments, as we would need to be for a monogenetic account of language emergence – especially catastrophic monogenesis. Nevertheless, features of language that evidence suggests are only expressed in literate and/or complex societies can reasonably be dismissed as candidates for selection at an early stage in human prehistory. Grammatical subordination seems to be one such (Kalmár 1985; Mithun 1984; Ong 1982; see Newmeyer 2002 and Wray & Grace *forthcoming* for interpretations of this evidence. Also see further discussion in section 4 below).

We are challenged, then, to accommodate in a model of language evolution the emergence not only of what humans *know* and what humans *do* with what they know, but also a mechanism by which one is mapped onto the other. This challenge cannot be reduced to a version of the old competence-performance debate – it is more fundamental than a simple failure of our production mechanisms to keep pace with some finely honed underlying system. Indeed, the notion that there is a fully-specified underlying system can no longer be taken for granted. Different theoretical approaches accommodate a lesser level of specification in different ways, but one feature that is increasingly playing a role is the question of how the experience of individuals can *support* a 'system' (more or less systematic) without *mastering* it. The implications are considerable for simulation research, since mastery of the system has always been the most difficult thing to model.

3.0 Mapping linguistic knowledge

3.1 Intuition and patterns of use

Establishing what humans 'know' about language is not easy. Corpus linguistics has opened our eyes to a great many things previously not recognised about patterns in linguistic behaviour, and these are surprisingly different from what our intuitions predict. Sinclair (1991) views linguistic intuition as "highly specific, and not at all a good guide to what actually happens when the same people actually use the language" (p.4). Much continues to be written, from both sides, about just why our intuitions don't match our language use. Increasingly, linguists are looking for a way to model our knowledge of language that can account for the mismatch in some plausible way. By plausible, I mean that there should be predictive power to the model of the relationship, rather than simply writing off one or the other component as mysterious or uninteresting. I shall propose such a model later. There are several things that it needs to accommodate.

On the performance side, one is the lexical patterns that corpus linguistics reveals: "Grammars based on intuitive data will imply more freedom of combination than is in fact possible" (Stubbs 1993: 17). Another is the disproportionate recurrence of certain formulations of common messages, where other formulations are also possible (Wray 2002a). A third is the absence of grammatical structures that are intrinsic to a basic grammatical theory. Meyer & Tao (2004) looked for gapping in the International Corpus of English (ICE). They found only 120 tokens in 17,629 examples of local coordination capable of supporting it (0.007%) (see also Favareau et al, in preparation).

On the competence side, we can note that our intuitions do not seem to match linguistic theory. Linguistics lecturers know that it is not always easy to convince a syntax class that the 'official' allocation of asterisks (on unacceptable sentences) is correct, even though it is supposed to reflect universal innate knowledge. Outside of the university setting, it is even more of an issue: Chipere (2000) found that relatively uneducated native speakers of English were very poor at making grammaticality judgements on complex sentences.

What does this signify? Do people 'know' things but have difficulty articulating their judgements? Do they have problems separating out strictly structural judgements from semantic and pragmatic variables? Or is it possible that they really don't *know* certain grammatical constraints? And if so, how should we define 'know' without raising questions of inegality at a fundamental level? Answering these questions impacts considerably on where simulation studies go in the next few years.

3.2 What does it mean to 'know' something?

Let us assume, with Newmeyer, that every human is equipped for Subjacency even if he or she is never called upon to engage with its constraints. Let us further assume that Subjacency did not arise as a response to some mental or communicative pressure that no longer exists (though this should be fully explored at some point). It follows that we have to explain our sensitivity to constraints such as those of Subjacency as a spandrel. This does not get us off the hook – a spandrel of *what*, and why did precisely this sort of constraint arise? Our model of how language evolved may be freed up by not having to build in certain constraints at the primary level, but it must still produce the constraints at the secondary level, and, if Newmeyer is correct, across the board, in humans that do and do not have occasion to apply it. In the same way, a model of how human physiology evolved does not need to build in, at the primary level, the human's capacity to ride a bicycle, since bicycle riding played no part in our physiological evolution. However, any evolutionary model that does not culminate in a modern physiology that includes the capacity to ride a bicycle is clearly wrong.

The question is, therefore, what sorts of models of language evolution can, without entailing the full gamut of linguistic realisations found today, nevertheless predict that they will be *possible* today? And how does one avoid building in unprincipled preadaptations, that is, shaping the model in certain ways in anticipation of what will be needed later?

4.0 Evaluating alternative models of linguistic knowledge

Once the door is opened, there are many paths to explore. In what follows, I pursue one particular line, which entails three potentially independent, but also linkable, insights into how we might account for the phylogenetic (and, in passing, ontogenetic) acquisition of linguistic knowledge.

4.1. Culturally based insights about language

Echoing Newmeyer, and in keeping with various observations made about the role of literacy in our perceptions of language (e.g. Givon 1979, Grace 1987, 2002a-c, 2003, Kalmár 1985, Ong 1982), Wray & Grace (*forthcoming*) propose that the extent to which humans engage with the free manipulation of linguistic structure is influenced by cultural and demographic variables. Drawing on ideas from Chafe (1985), Kay (1977), Laycock (1979), Mithun (1984), Thurston (1987, 1989) and Trudgill (1989, 2002), we suggest that the default level at which humans exploit the creative potential of language is somewhat conservative. However, under certain sorts of conditions, such as prolonged contact with adult learners of the language, and social structures that create and sustain linguistically defined in- and out-groups, additional engagement with the mechanics of language becomes necessary. This augmented engagement serves explicitness and form-meaning predictability. Messages will become easier to interpret out of their temporal and/or sociocultural context, and it will be easier to create and understand novel messages[2].

According to this model, there is a fluid range of potential experience for individuals in relation to the manipulation of their language. Those operating at the default level are those whose environment and way of life feature a large measure predictability in daily behaviour and message content, such that their ability to understand, and produce, novel messages is rarely challenged to handle more than limited lexical variation within much-used message frames.

Under different sociocultural conditions, however, these same individuals will increase their engagement with the mechanisms of their language, to accommodate the greater need for explicitness in messages. To take a modern-day example, a specialist 'in-group' such as car mechanics may possess a highly contextualised code for discussing their work amongst themselves. This code will feature jargon that affords short-cuts in relation to customary reference – jargon that must be acquired as part of an apprenticeship, both for social and practical reasons. The code may simultaneously facilitate in-group communication, furnish its users with a means of preventing outsiders from understanding certain secret messages, and signal, through that exclusion, professional identity. However, the code is of little use when the mechanic has to discuss the state of a car with its non-specialist owner. Jargon must be glossed, and technical procedures explained rather than just named. In order to achieve an adequate level of effective communication, both mechanic and customer must have access to a code of common terms and ways of formulating them. This *lingua franca* will be most effective if its components are subject to transparency and regularity, so that meaning can be teased out where it is not available outright. For practical reasons, the *lingua franca* will also need to be transferable to other similar situations.

Under this account, individuals' linguistic capabilities, in terms of novelty of expression, are conservative by default, but easily stretched when necessary. This 'elastic' model can explain the contrast between what we evidently *can* do, and what we customarily *do* do. Unlike models that assume all individuals to be exercising the full range of linguistic manipulation all the time, there is no need here to account for the failure of that full potential to be realised in most instances of language in use. Rather, the potent variable is that which forces the individual, temporarily or more permanently, out of the default mode in which messages are expressed the same way as last time, and in which the interpretation of input is filtered through pragmatic and contextual predictability first, and only further analysed if necessary.

4.2 Processing by need, not principle

The processing mechanism upon which I propose the elasticity to be based is *needs only analysis* (NOA) (Wray 2002a: 130-2), and I shall explain it first in relation to child language acquisition. The principle of NOA is simply that the individual does not break down input any further than is necessary to extract or create meaning. That is, there is no gratuitous analysis of form beyond the point where form-meaning mapping is sufficient for the present comprehension event, or for the construction of the presently required output. Over a period of time, an accumulation of event-specific comprehension and production requirements will lead to the identification of many small, recombinable items, plus rules for their combination. However, large units that never require such reduction will remain intact, and this will result in a mixed inventory of small and large items, as determined by the patterns in the input. In many cases a large unit will loosen up to permit limited morphological or lexical variation, creating a partially lexicalised frame (e.g. *at the end of the __; NP1 pull + TENSE/AGR NP2+ POSSESSIVE leg*).

Once a reliable meaning can be associated with a particular unit, that unit will be subject to privileged retrieval for that meaning, both in comprehension and production. As a result, in comprehension, even if the unit is internally complex (e.g. *don't count your chickens; the thing is)* other logical but disfavoured

[2] It should not be inferred that there is no scope for novelty without this augmentation. The default level is above the threshold at which new meaning can be expressed and understood. However, more adventurous combinations will, indeed, be possible with augmented engagement. How this works is further described later.

interpretations of the unit will not be countenanced. Thus, under normal circumstances, *tear along the dotted line* will not be interpreted as an instruction to sprint along the middle of the road because another meaning is already associated with that complex unit. In production, other entirely grammatical formulations of the intended message will tend not to be produced (e.g. *rip along the marked pathway*), because it is easier to retrieve a preformed unit than to construct one[3].

In further illustrating how NOA works, it is useful to look at the extreme end. An idiom like *by and large* has reliable meaning at this three-word level, and so there is no impetus to break it down any further. As a result, we do not construe the meaning of 'large' with reference to its occurrence in this idiom, and we are not tempted to create novel meanings by changing items within it, e.g. **by but large; *by and small*.

A slightly less fixed idiom like *from now on* will also be initially assigned a holistic meaning, but over time, input will reveal the potential to loosen the fixedness of *now* to permit certain other time-related items, e.g. *from then on, from Tuesday on, from that moment on*, but without loosening the constraints on the first and third words, since no evidence will arise from input to suggest that **till now on* or **from now off* are possible. It follows that intuitions about what 'sounds right' are closely attuned to experience of input and how that has affected the specification of looseness for a given lexical configuration.

Literate language users will encounter a broader range of messages, many decontextualised and explicit. As outlined earlier, explicitness is achieved by virtue of predictability and system. Learning to be literate is, as any teacher will verify, much more than learning to recognise and form letters on the page. It entails the mastery of the manipulation of language in the service of explicitness. What we read counts as input, and the desire to read and understand changes the 'needs' dimension of needs only analysis. We need to analyse more in order to cope with the new ideas and structures we encounter, and this influences what we know of the language and what we can do with it. In addition, general western-style education will invite us to perceive the whole as a product of its parts, further influencing our belief that the way to master language is to master its smallest components. In adult language learners, this compulsion is, I have elsewhere concluded, virtually irresistible, even where it specifically interferes with effective communication (Wray 2004). In contrast, NOA predicts that children, being preliterate and uneducated, will overlook even quite explicit linguistic regularities, if those patterns do not map onto something they need for the extraction of meaning. In line with this prediction, Bergen (2001) found that child native speakers of Esperanto introduced irregularities into the perfectly regular system. This is something that would not be predicted by a model of language acquisition that entailed the principled pursuit of fully systematic patterns (see Wray & Grace *forthcoming* for a full exploration of these issues).

4.3 Cultural inheritance

Needs only analysis naturally entails that we can inherit linguistic material that we have not ourselves analysed. This is comparable to being able to drive a car that you couldn't build or fix, though you might be able to dabble with more or fewer components, such as filling up the water reservoir, putting air in the tyres, replacing a light bulb, or changing the spark plugs. The analogy is useful here, since our ability to do such maintenance jobs on our cars is also determined by need. Leaving aside going to car maintenance classes or training as a mechanic – the equivalent of taking language classes in school – one will tend not learn how to change the plugs until the plugs need changing, and only then if there is no alternative to doing it oneself. In the same way, the extent to which we manipulate language depends on the situations we find ourselves in, and what the alternatives are. In ordinary educational settings we may learn to manipulate the language sufficiently to recognise and command different styles for different purposes. In a poetry class we would learn to analyse and create novel formulations that push at the boundaries of customary semantic (and in some cases grammatical) practice. A professional writer would learn to hone text to create subtle effects. And, at the most extreme end, someone who does cryptic crosswords crosses the boundary into the bizarre misanalysis of what appear to be normal linguistic formulations. This would be the equivalent of melting down a car tyre to extract the chemical components, perhaps.

That we can also duck out of analysis if the opportunity exists is exemplified by Rehbein's (1987) observation that Turkish guest workers in Germany did not raise their linguistic skills to meet their communicative need, but rather curtailed their communicative need (by avoiding certain situations) to match their existing linguistic skills. Many more people demonstrate the same choice when, on holiday in a foreign country, they elect self-service shopping over counter- or market-stall transactions.

One particular consequence of NOA is of significance here. It relates to the potential for the inherited

[3] NOA explains why our language use is different from our linguistic insight. The former is subject to NOA, while the latter can be marshalled to judgements about what is logically possible too. NOA can also explain why our judgements about what words mean and what they can combine with do not match how we use them (see Wray 2002a: 276-7).

language to carry material that is not – and in due course could not come – under the control of its current users. In what we might term the 'local' consequence, an originally regular formulation will, if not subject to paradigmatic variation in the input – and this might be for no other reason than that the messages expressed by that paradigmatic variation are never needed – fail to be analysed. Where users do not activate the compositional structure of a word-string, they will be less likely to update it in the light of diachronic or other language change, or to correct phonologically conditioned errors. Additionally, there may be semantic drift as the meaning is no longer grounded through its component parts. Thus, over time, idiomatic expressions can become stranded as fossils, maintaining lexical and/or grammatical forms that are no longer active in the language (e.g. *curry favour; rather thee than me; director general*), or phonological indeterminacy (e.g. *streaks/streets ahead; off his own back/bat*). The more irregular an item becomes, the more it resists analysis, until the individual encountering it has no choice but to accept it holistically and assign a global meaning to it (e.g. *by and large*).

4.4. Corollaries for modelling language evolution

From this local consequence, longer-term scenarios become logically possible, and these have a direct bearing in how language processing (in evolution or acquisition) is modelled. For clarity, it must be noted that while NOA is proposed to be an identical process in both the phylogenetic and ontogenetic acquisition of language, there is, of course a key difference, in that children apply NOA to input deriving from an existing linguistic system (albeit not a fully specified one), while language itself either emerged spontaneously, or evolved from something that was not itself language. The mechanisms by which language is acquired by the individual have been adequately covered above, so we shall focus now on NOA in language evolution.

One scenario for how irregularities arise under a rule-based system requires us to envisage some parent language starting out perfectly formed, with no irregularities. Although implausible in all but certain kinds of catastrophic account, this scenario is worth briefly considering because a good test of NOA would be the modelling of its effects on regular initial input. In models of linguistic knowledge that assume humans to seek a fully specified grammar that will work optimally on fully regular input, it is difficult to explain why, if a language was, at some point in the past, fully regular, it should ever have ceased to be so. In contrast, as demonstrated in sections 4.2 and 4.3 above, under NOA irregularity is predicted to emerge, even if the starting place is one of perfect regularity (Bergen 2001).

However, there may never have been perfect regularity from which human languages have strayed. An alternative scenario is that language has always been irregular, either entirely or at least at the edges. Various explanations might be offered for this, including some vagary in the fundamental operations of the human mind on linguistic material, and the accumulation of complexity under the gradual emergence of fully formed language out of precursor systems. NOA is consistent with all such scenarios, but also with one in which the building blocks of the languages we manipulate today are the – albeit much changed – descendents of forms extracted through *post hoc* rationalisation from a phonetically expressed holistic protolanguage (Wray 1998, 2000, 2002b)[4].

In this account, the emergence of language is preceded by the capacity to use 'large words' – that is, holistic sound- or gesture-strings – to express common manipulative messages. These messages might reasonably be envisaged to be more numerous than those in other animals, and to be the product of more intricate and controlled articulatory movements. But they are still holistic units associated with specific messages used to signal within a narrow functional range. Also pre-specified is semanticity, in the specific sense of a capacity to discern and interpret things in the world. This seems uncontroversial, since (a) such discernment is fundamental for survival in numerous species, and (b) there is no implication that what is discerned can be labelled, nor that individuals would, if they could label, share judgements about what to label (compare Steels & Kaplan 2002 experimenting with AIBO). The holistic units, of course, must have their own semantic representations, but, as with animal calls, these are grounded in observable action[5] and if not so-grounded they will cease to be viable. This give the process of NOA, when it kicks in[6], the scope to itself fuel the emergence of semantic categories, based mostly on existing perceptual preferences but partly on pure chance.

In modern first language acquisition, variation in the input is a product of the rule-based flexibility in a system that is being manipulated by those who al-

[4] The holistic protolanguage story as I have framed it to date has focussed on a phonetic realisation. However, as Arbib (2005) demonstrates, a gestured protolanguage is also plausible and, I think, presents no major problems to my overall thesis.

[5] Animals seem to vary in the extent to which their holistic cries for basic communicative functions are innate versus learned. In the case of the precursor of human language, the signals must, of course, be learned. This in itself will narrowly constrain the messages to certain kinds of groundable meanings.

[6] Kirby's (2001) bottleneck account goes some way to providing a rationale for the use of NOA, though we cannot avoid the need to explain how our ancestors became equipped to do so, nor why they had not applied analysis before.

ready command it. As a result, the child has many clues about what can change and what can't and what the effect is on meaning. In contrast, in the evolutionary context, the dividing up of holistic forms with complex meanings will be down to coincidental repeated mappings of sub-forms and sub-meanings, supported by pragmatically-motivated contingencies such as hyper-correction and semantic fission[7] (Wray 1998, 2000, 2002b).

So much for words. What of grammar, though? Why is language grammar the way it is? Neither NOA theory not the holistic protolanguage model as I have framed it to date directly explains why grammatical patterns are what they are. However, what we can now do is separate out the origination of a system from its perpetuation. NOA in child language acquisition acts on an existing system. It maintains and perpetuates the parts that work. Anything that has no productivity is not analysed and so ends up on the irregular periphery. NOA acts, in fact, as a filter that ensures a good mapping between what can be said and what needs to be said, while maintaining the edge necessary to meet unexpected additional demands. Since the system extrapolated by an individual on the basis of NOA is a product of (a) mapping the material of actual input onto the communicative need of the individual and (b) the principle of economy (that is the principle of linguistic analysis according to need), it also follows that the active grammar can be subject to reformulation in the minds of individuals, something that easily accounts for phenomena such as grammaticalization. Because there is no underlying template to demand that exemplars of this or that be found, such changes in a language's make up are, indeed, much easier to account for under NOA than in theories that attribute to humans a fully specified linguistic system.

This may explain how grammar is perpetuated and managed once it exists, but one must still ask how grammar arose. Once it is there, pass the parcel is all very well. But how did the parcel get wrapped up in the first place? If Newmeyer and others are right, we can exclude certain features of modern language grammars from consideration, because they are only encoded in response to later cultural conditions. But whatever a child can work out in the early years of acquisition must surely be fundamental to how the human mind works, and must have played a role in why all human languages have turned out to have certain common properties. NOA in language evolution places the burden entirely on independently evolved cognitive mechanisms, exapted to find a way of representing not only entities, properties and actions, but also relationships between them. The NOA account does not require latent specific language capabilities (though it can accommodate them). All it needs is independently evolved mechanisms for perception, thought and memory management. Whether such mechanisms can really account for language grammar is a question that many linguists who favour the 'general cognitive mechanisms' approach are presently engaged in answering.

4.5. Nicaraguan Sign Language and NOA

Since the proposed scenario considerably downplays the role of an underlying drive to find linguistic structure for its own sake, it is worth considering the issue of creolisation, and in particular the case of Nicaraguan Sign Language (NSL) (e.g. Kegl 1994; Senghas 1995; Senghas & Coppola 2001). A favoured interpretation of the progression of NSL is that, as Bickerton (1988) suggests, children look for evidence of realisations of UG in their input, and if they fail to find any (as with pidgin input) they will impose UG, using default settings. Whilst it is still premature to judge whether the NSL data supports Bickerton's proposal, it does seem to me that there are other factors involved, and that the emerging patterns are, so far, equally consistent with NOA (Wade 2004). As outlined earlier, NOA operates in response to communicative need. In fully formed languages the input of adults offers ample guidance for the child to establish the points of variation and the tendencies to fixedness that are characteristic of the language in use in the speech community. That is, the language can be relied on to furnish a means of expressing key concepts and relations that the human brain perceives and that the individual will want to use language to articulate. Beyond that, the child will presumably develop its preferences for semantic differentiation in tandem with the development of awareness of how its language varies to convey different meanings. Furthermore, the child may tend to trust that it will not be confronted by the need to say something that the adults around it cannot also say – a supposition which if not true in the first instance will soon become so.

In the case of an emerging language, however, things are different. If the child is unable to access an effective means for expressing basic concepts and

[7] That is, pragmatics (or expediency) might fuel an assumption that two not quite identical forms that apparently mean the same probably are supposed to be identical (hypercorrection). In another case, it might be concluded that two non-identical forms apparently meaning the same must in fact mean different, but related things (semantic fission, or 'the splitting of the semantic space'). The first will precipitate the fixing of form-meaning pairs while the latter provides an impetus for the creation of new meanings to match forms, e.g. hyper- and hyponyms, restricted collocations etc. For instance if forms X and Y are both thought to mean, say, 'not fit to eat', the context may be over-interpreted, so that X is taken to be the term used with food A while Y is used with food B. If A and B have different properties, what begins as a semantically-neutral collocational restriction will easily lead to new nuances of meaning, as we see in English with *rancid butter, sour milk, over-ripe cheese*. Modern examples are not an adequate parallel for *ab initio* semantic fission but they do indicate that we can handle it.

relations, then either the child must remain mute (e.g. Schaller 1995) or it must establish a means of making good the shortfall of expressive material. NOA will be sufficient to propel the child into looking for an opportunity to extract manipulable parts, and relationships between them, from anything it has access to, be it spoken pidgin, home sign or an emerging language (Wade 2004). That is, communicative need will, if able to, drive the establishment of sufficient material to create exactly those meanings that are needed (and by default, but not design, some others).

However, it must be recognised that the process of establishing a database of such manipulable forms and structures will certainly be influenced and precipitated by any independently developing awareness on the part of the child of what language has the potential to do, such as would arise under exposure to secondary, culturally-based input – literacy, general education – or to a fully-fledged language. Although NSL evolved in the playground and school bus, it did so precisely because communication the classroom was not based around signing, but around Spanish. That is, we should not overlook the role that the classroom might have played in determining the needs of the children to express certain kinds of messages in certain kinds of ways. In short, it is reasonable to conjecture that NSL has been – and continues to be – shaped by other kinds of input than the signing itself. The effect might be, generally, to guide the language towards the development of particular kinds of features, and/or to engage Spanish, or other languages in the environment such as ASL, as substrates.

5.0 Conclusion: the potential of the short cut approach

Needs only analysis predicts that individuals might appear to command a range of complex linguistic functions while actually not having full command of them – though they would retain the capability to develop a fuller command of them should the need arise. Insofar as configurations can used before they are analysed (if they ever are), it remains possible that some – particularly those that require an arcane rule to generate them – might not be under the generative control of anyone other than linguists and pedants (Grace 2002a-c, 2003). This is in line with our undoubted ability to use foreign phrases appropriately without having a command of the grammar underlying them. However, it potentially extends well beyond this, to, at the extreme end, the possibility that judgements about relationships between elements (or gaps) in embedded clauses are not innately specified, but rather are developed in the course of cultural linguistic training, on the basis of institutionalised *post hoc* rationalisation.

Two key component of the NOA model can be separated out in the context of computational systems. The first is the handling of unanalysed chunks. That a surprising amount of effective communication can be achieved entirely formulaically has been demonstrated by artificial systems such as TALK – a communication aid for non-speaking people (Todman et al 1999a,b; Wray 2002c) and TESSA – a limited English-BSL translation system for the British Post Office (Wray et al 2004). Both of these are based on the handling of predicable, holistically material, and each has a limited capacity to generate new messages, either through on-line editing (TALK) or partially lexicalised frames (TESSA). However, neither system has the second element of NOA: a dynamic learning component.

NOA with the dynamic component is most closely exemplified in the modelling of Dominey, Vogt, Kirby and others. Kirby (2001) found that holistic input will, when broken down, not necessarily resolve everything into consistent unit types or sizes. Islands of non-compositional material will remain. It might be tempting for such modellers to aim to find a set of parameters that removes this ‘problem’, since we have come to expect that full systematicity is the goal. But NOA suggests that this is not necessary. Not only does the input not need to be structured in order for structure to be extracted; neither does the output have to be fully structured for the process to be in some way representative of what humans do with language. An NOA based model will be adept at dealing with novel strings (albeit only after catch-up analysis), but, usefully, it will not overgenerate. Overgeneration is a consequence of over-specification relative to the target model. It is time that we expected a good model of language to be able to simultaneously handle novelty and predict the shortfall between what it is theoretically possible to say and what we actually do say (Pawley & Syder 1983).

Needs only analysis offers an alternative approach, which could have far-reaching consequences for language-focussed AI research. NOA minimises unnecessary actions, but enables the system to extract additional components when it needs to. A system that is fed a plausible approximation of contextualised normal human language will never need to identify the full complement of potential atomic units, assign them to categories and generalise about their potential to appear elsewhere. That is, the system will remain underspecified, and in that regard, unstable – or, to put it more positively, continually open to modification in response to new evidence. Such dynamism is attractive as a basic characteristic in a model of human behaviour.

As I hope I have shown, under NOA, the modelling of language acquisition and language evolution are essentially the same thing. Ontogeny does, in this case, recapitulate phylogeny (other than that in evolution there was no system to find, whereas today there is). Since input will vary in response to the system possessed by those that produce it, it follows that language emerged in the species somewhat more gradually than it emerges in the modern child. However, that might well mean a few generations rather than thousands, and we can still reasonably construe the emergence of modern human language in our species as a relatively swift event. Significantly, we are not bound to propose that language arose, and then everything became stable. NOA offers a plausible vehicle for the continuing directional changes that Newmeyer (2002) reviews. It also draws in the secondary variables of culture and social organisation, to also play their part in the continuing evolution of language.

References

Arbib, M. A. 2005. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* [in press].

Bergen, B.K. 2001. Nativization processes in L1 Esperanto. *Journal of Child Language* 28: 575-595.

Bickerton, D. 1988. Creole languages and the bioprogram. In Newmeyer, F. (ed.) *Linguistics: the Cambridge Survey. II: Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press, p.268-284.

Chafe, W. 1985. Linguistic differences produced by differences between speaking and writing. In Olson, D.R., Torrance, N. & Hildyard, A. (eds.) *Literacy, language and learning*. Cambridge: Cambridge University Press, 105-123.

Chipere, N. 2000. *Processing embedding in humans and connectionist models.* Unpublished PhD Thesis, University of Cambridge.

Favereau, D., Goodwin, C., Lee, N., Mikesell, L, Schumann, J.H., Tao, H., Veronique, D. & Wray, A. in preparation. Language Evolution: What Evolved? Position paper following Language Learning Round Table, UCLA, Nov.04.

Givon, T. 1979. *On understanding grammar*. New York: Academic Press.

Grace, G.W. 1987. *The linguistic construction of reality*. London: Croom Helm.

Grace, G. W. 2002a. Collateral damage from linguistics? 1: The post-Chomskyan paradigm and its underlying assumptions. *Ethnolinguistic Notes* 4, 21. http://www2.hawaii.edu/~grace/elniv21.html

Grace, G. W. 2002b. Collateral damage from linguistics? 2: the cultural evolution of language. *Ethnolinguistic Notes* 4, 22. http://www2.hawaii.edu/~grace/elniv22.html

Grace, G. W. 2002c. Collateral damage from linguistics? 3: The role of culture-centrism. *Ethnolinguistic Notes* 4, 23. http://www2.hawaii.edu/~grace/elniv23.html

Grace, G. W. 2003. Collateral damage from linguistics? 4: Do we really know what kind of language the language acquisition device prepares us to acquire? *Ethnolinguistic Notes* 4, 21. http://www2.hawaii.edu/~grace/elniv24.html

Kalmár, I. 1985. Are there really no primitive languages? In Olson, D.R., Torrance, N. & Hildyard, A. (eds.) *Literacy, language and learning*. Cambridge: Cambridge University Press, 148-166.

Kay, P. 1977. Language evolution and speech style. In Blount, B.G. & Sanches, M. (eds.) *Sociocultural dimensions of language change*. New York: Ac. Press, 21-33.

Kegl, J. 1994. The Nicaraguan Sign Language Project: an overview. *Signpost* 7 (1): 24-31

Kirby, S. 2001. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation* 5 (2): 102-110.

Kirby, S. 2002. Learning, bottlenecks and the evolution of recursive syntax. In Briscoe, T. (ed.) *Lingisitic evolution through language acquisition: formal and computational models*. Cambridge: Cambridge University Press.

Laycock, D. 1979. Multilingualism: linguistic boundaries and unsolved problems in Papua New Guinea. In Wurm, S.A. (ed.) *New Guinea and neighboring areas: a sociolinguistic laboratory*. The Hague: Mouton, 81-99.

Meyer, Charles, and Tao, Hongyin. 2004. Gapped Coordination Constructions in English Discourse and Grammar. Paper presented at the Fifth North American Symposium on Corpus Linguistics, Montclair State University, New Jersey. May 21-23, 2004.

Mithun, M. 1984. How to avoid subordination. *Proceedings of the tenth annual meeting of the Berkeley Linguistics Society,* 493-523.

Newmeyer, F.J. 2002. Uniformitarian assumptions and language evolution research. In Wray, A. (ed.) *The transition to language*. Oxford: Oxford Univ. Press, 359-375.

Ong, W.J. 1982. *Orality and literacy: the technologizing of the word*. London: Methuen.

Pawley, A. & Syder, F. H. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J.C. &. Schmidt, R.W.(eds.) *Language and Communication.* New York: Longman, 191-226.

Rehbein, J. 1987. Multiple formulae: aspects of Turkish migrant workers' German in intercultural communication. In Knapp, K., Enninger, W. & Knapp-Potthoff, A. (eds.) *AnalysingIntercultural Communication*. Berlin: Mouton, 215-248.

Schaller, S. 1991. *A man without words*. Berkeley, CA: University of California Press.

Senghas, A. 1995. Children's contribution to the birth of Nicaraguan Sign Language. PhD thesis, MIT; MIT Working Papers in Linguistics, No. SENG01.

Senghas, A. & Coppola, M. 2001. Children creating language: how Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science* 12 (4): 323-328. http://icg.harvard.edu/~psy2020/Snedeker_Readings/Senghas-and-Coppola.pdf

Sinclair, J. McH.1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Steels, L. & Kaplan, F. 2002. AIBO's first words. *Evolution of Communication* 4 (1): 3-32.

Stubbs, M. 1993. British traditions in text analysis: from Firth to Sinclair. In Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.) *Text and Technology: in honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 1-33.

Thurston, W. 1987. Processes of change in the languages of north-western New Britain. *Pacific Linguistics B99*. Canberra: The Australian National University.

Thurston, W.R. 1989. How exoteric languages build a lexicon: esoterogeny in West New Britain. In Harlow, R. & Hooper, R. (eds.) *VICAL 1: Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics*, Auckland, New Zealand, January 1988. Auckland, NZ: Linguistic Society of New Zealand, 555-579.

Todman, J., Rankin, D. and File, P. 1999a. The use of stored text in computer-aided conversation: a single-case experiment. *Journal of Language and Social Psychology* 18 (3): 320-342.

Todman, J., Rankin, D. and File, P. 1999b Enjoyment and perceived competence in computer-aided conversations with new and familiar partners. *International Journal of Rehabilitation Research* 22: 153-154.

Trudgill, P. 1989. Contact and isolation in linguistic change. In Breivik, L.E. & Jahr, E.H. (eds.) *Language change: contributions to the study of its causes*. Berlin: Mouton de Gruyter, 227-237.

Trudgill, P. 2002. Linguistic and social typology. In Chambers, J.K., Schilling-Estes, N. & Trudgill, P. (eds.) *handbook of lingusitic variation and change*. Oxford: Blackwell, 707-728.

Wade, N. 2004. Deaf children's *ad hoc* language evolves and instructs. *New York Times* Sept 24th

Wray, A. 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18: 47-67.

Wray, A. 2000. Holistic utterances in protolanguage: the link from primates to humans. In Knight, C., Studdert-Kennedy, M. & Hurford, J. (eds.) *The evolutionary emergence of language: social function and the origins of linguistic form*. Cambridge: Cambridge University Press, 285-302.

Wray, A. 2002a. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. 2002b. Dual processing in protolanguage: performance without competence. In Wray, A. (ed.) *The transition to language*. Oxford: Oxford Univ. Press, 113-137.

Wray, A. 2002c. Formulaic language in computer-supported communication: theory meets reality. *Language Awareness* 11 (2): 114-131.

Wray, A. 2004. 'Here's one I prepared earlier': formulaic language learning on television. In Schmitt, N. (ed.). *The acquisition and use of formulaic sequences*. Amsterdam: John Benjamins, 249-268.

Wray, A., Cox, S., Lincoln, M. & Tryggvason, J. 2004. A formulaic approach to translation at the post office: reading the signs. *Language and Communication* 24: 59-75.

Wray, A. & Grace, G.W. forthcoming. The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*.