Department of Linguistics

Stockholm
University

Proceedings

PERILUS 2011

Symposium on Language Acquisition and Language Evolution

The Royal Swedish Academy of Sciences
December 1-2, 2011

Stockholms
universitet

Symposium on
Language Acquisition and Language Evolution

Kungliga Vetenskapsakademien och Institutionen för lingvistik, Stockholms universitet

1-2 december 2011

Beijersalen, Kungliga Vetenskapsakademien, Stockholm

Symposiet hålls på engelska

Symposium on Language Acquisition and Language Evolution

Beijersalen, Kungliga Vetenskapsakademien, Stockholm

Program

1 december 2011

Tidsram 13.00 – 17.00

Öppningsanförande: Björn Lindblom
Professor emeritus, Institutionen för lingvistik, Stockholms universitet

Stephen Freeland, astrobiologi, NASA, Hawaii, USA

Pierre-Yves Oudeyer, robotik, INRIA Bordeaux; Frankrike

Peter MacNeilage, professor emeritus, psykologi, University of Texas at Austin, USA

15.00 Kaffe

Björn Lindblom, professor emeritus, Institutionen för lingvistik

2 december 2011

Tidsram 09.00-13.00

Diskussion

Jan Anward *Units from interaction*
Professor, Linköpings universitet

Per Linell *Rethinking Language*
Professor, Institutionen för språk och kultur, Linköpings universitet

10.30 Kaffe
Peter Pagin
Professor, Filosofiska institutionen, Stockholms universitet

Magnus Enquist
Professor, Centrum för evolutionär kulturforskning, Stockholms universitet

Avslutande populärvetenskaplig debatt 12.00–13.00

Moderator: Patrik Hadenius
Chefredaktör på Forskning & Framsteg, Språktidningen samt Modern Psykologi

Institutionen för lingvistik

Stockholms universitet 106 91 Stockholm
Telefon 08-16 23 47

Besöksadress Universitetsvägen 10 C
E-post:exp@ling.su.se www.ling.su.se

Stockholms
universitet

Table of contents

Preface

This volume contains articles based on the authors' contributions to the Symposium on Language Acquisition and Language Evolution which was held at The Royal Swedish Academy of Sciences in Stockholm on the 1st and 2nd of December 2011, by initiative and funding from the Department of Linguistics at Stockholm University.

The symposium was intended as an opportunity for scientists from different research areas to interact and discuss complex dynamic systems in relation to the general theme of "Language Acquisition and Language Evolution". Complex dynamic systems are characterised by hierarchical and combinatorial structures that can be found in quite different scientific domains. From a broad perspective, there are general parallels in the way human language, biological organisms and ecological systems are organised and the symposium aimed at discussing those issues from an interdisciplinary point of view.

The goals were largely achieved. The invited speakers were a small group of top researchers from Phonetics, Linguistics, Biology, Philosophy, Astronomy and Computer Sciences who contributed with very inspiring and insightful presentations that effectively stimulated fruitful discussions among all the participants. The articles in this volume represent the themes brought up by the speakers and capture the essential aspects raised by the speakers. Not surprising, however, the articles cannot do justice to the climate of spontaneous scientific debates and public involvement that the corresponding live presentations generated. They also fall short from conveying the glowing engagement of the class of interested and ambitious high-school students from Kungsholmens Gymnasium in Stockholm, who participated in the closing session of the symposium. Under the guidance of the science journalist Patrik Hadenius, the eagerness and interest of these high-school students flourished rapidly and their commitment to learn about scientific issues became one of the most memorable highlights of the symposium. It was extremely rewarding for all the speakers and enhanced even further the enthusiasm of the participating scientists. In the course of this friendly and intensive debate of scientific issues, the speakers found themselves literally surrounded by the high-school students in their pursuit to learn more and were extremely happy about that overwhelming interest!

We are proud to publish in this volume these important contributions on the broad theme of Language Acquisition and Language Evolution.

Stockholm, August 2013

Francisco Lacerda
francisco.lacerda@ling.su.se

The units of speech – A bio-cultural perspective

Björn Lindblom

Department of Linguistics, Stockholm University

lindblom@ling.su.se

Abstract

Humans have large vocabularies and learn new words all of their lives. From infancy to adolescence, children acquire lexical entries at a remarkably fast rate. These accomplishments are linked to the uniquely human method of coding information: the combinatorial use of discrete entities such as the phoneme. This method creates the conditions for open-ended lexical systems thereby contributing to the singular expressive power of human languages. After reviewing a broad range of experimental evidence on speech and non-speech we conclude that a strong case can be made for the claim that the hallmark of true phonology - the combinatorial structure of sound patterns – has behavioural origins.

Keywords: vocabulary learning; world's sound systems; perceptual contrast; size principle; syllable; phonotactic patterns; coarticulation; discrete units; positional control; motor equivalence; kinematic constancies; cognitive growth; vocal imitation; recombination; origin of phonemes.

Goals

Human language is a product of biology and culture. This truism is readily acknowledged in contemporary linguistics and forms an informal part of the ongoing academic narrative about language. However, when it comes to the more ambitious task of proposing formalized and explanatory theories of language, the bio-cultural perspective is curiously bypassed, even explicitly down-played. A case in point is the influential agenda persuasively promoted by Chomsky and embraced by many leading linguists. Anderson's "Why phonology isn't 'natural"" is a well-known attempt to spell out what this program entails for the study of sound patterns (1981).

There are several reasons for the lack of interest in language as behaviour in linguistics. An important one is the fact that the field continues to be under the sway of 'structuralism' which has been, and still is, firmly anchored in the belief that *'the classical Saussurean assumption of the logical priority of the study of langue (and the generative grammars that describe it) seems quite inescapable"* (Chomsky 1964:52).

Undeniably, this school of thought has been a huge academic success sociologically thereby significantly raising the scientific image of the discipline.

The focus of the present contribution is on the units of speech such as syllables and phonemes. The limitations of the structuralist approach become abundantly clear when we ask: "Where do these units come from?" In the prevailing paradigm that is an unasked question. Nonetheless, with respect to the phoneme, there is a possible answer: "It comes from phonemic analysis". In other words, it comes from observed linguistic facts filtered through an operational procedure aimed at identifying phonetic differences that change meanings. The output of the method is an abstract unit stripped of its intrinsic phonetic content and specified by a set of attributes or ‘features’ (as derived from the phonetic differences).

The shortcomings of the ruling paradigm are also exposed in the context of applications. Knowledge about language is needed in many areas facing complex educational, clinical and technological problems. Those problems are more likely to be linked to aspects of language use (language learning and teaching, speech and language pathologies, man-machine interaction etc.) than to language as an ‘autonomous’ (performance-independent) object.

The goal of the present paper is to demonstrate the feasibility of a non-structuralist approach to the basic units of speech. Accordingly we reverse structuralist priorities starting from substance (performance-based factors) in a quest for possible behavioural causal factors underlying sound structure formation.

Some facts in need of explanation

Human vocabularies are large and open-ended

Measurements of vocabulary size provide an important datum about the language learning process.

Figure 1 presents findings from studies of American English and Swedish children (Miller 1977, 1991 and Edlund 1957). Straight lines were used to connect the points obtained by pooling the two data sets. The resulting pattern is s-shaped suggesting that learning has its most rapid course between 8 and 12 and then slows down and reaches 60,000 root words at 17 years - a conservative estimate for an average America high school graduate (Miller 1991).

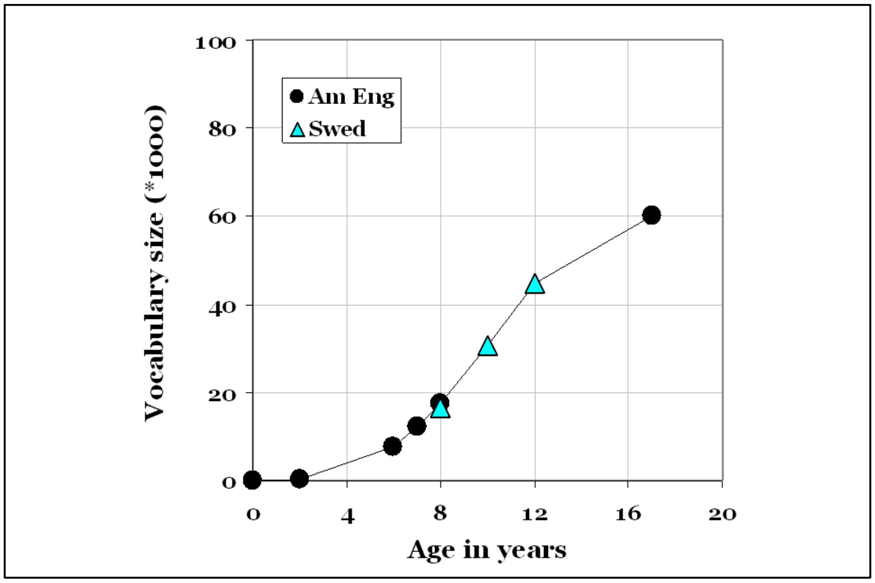


Figure 1 Average American English and Swedish data on vocabulary growth as a function of age.

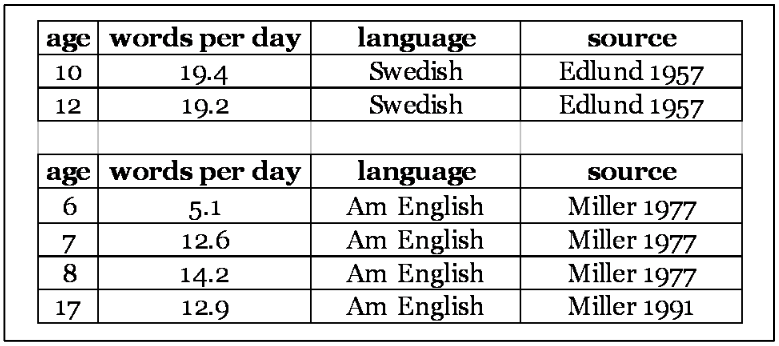
This observation is in agreement with a well-known fact: Word learning continues throughout adult life. The learning curve would thus be expected to continue to rise above the 60,000 mark after the 17th birthday, albeit more slowly. ‘Open’ word classes, such as nouns, adjectives and verbs continue to receive new input, whereas ‘closed’ classes, grammatical morphemes, tend to remain fixed in size (Teleman 2012).

Vocabulary growth is open-ended. As word formation rules remain productive, there is no fixed upper limit.

Word learning is fast and unsupervised

To shed some light on numbers like those in Figure 1, Miller chose to describe his findings in terms of learning rate as measured in 'words per day'. We repeated this exercise for the data of Figure 1 by using adjacent observation points. For each pair we divided the increase in number of words by the number of days elapsed. For instance, from 8 to 10 years old there is an increase from 16630 to 30810 Swedish words. The average number of words learned per day then becomes (30810-16630)/((10-8)*365)= 19.4.

Table 1 Rate of word learning expressed in words per day.



age	words per day	language	source
10	19.4	Swedish	Edlund 1957
12	19.2	Swedish	Edlund 1957

age	words per day	language	source
6	5.1	Am English	Miller 1977
7	12.6	Am English	Miller 1977
8	14.2	Am English	Miller 1977
17	12.9	Am English	Miller 1991

In reflecting on these numbers we should bear in mind that they are averages. They are not to be taken literally. They do not imply that learning say ten words takes only one day. Words probably vary with respect to how long it takes to learn them. Initially there may be a large heterogeneous set in which individual items are at more or less advanced stages, some requiring more practice and exposure, others on the brink of complete mastery. Then an organizing principle emerges and, from then on, the words-per-day measure starts to return large average numbers.

Having said that, we still have reason to marvel at the speed at which this development unfolds.

The combinatorial nature of lexical packing

In short, vocabularies are large and open-ended. Normal children learn them fast and without explicit teaching. How do they do that?

Part of an answer must be linked to the fact that, unlike animal communication systems, the contents of human vocabularies is not a set of holistic Gestalt patterns but is built from a small number of discrete units (phonemes) that are combined to form bigger units such as syllables, morphemes and words. Combinatorial coding is an extremely efficient method for specifying large amounts of information and the organizing principle of lexical inventories is the recombination of phonemes.

Phonemic coding is one of the keys to the size of human vocabularies.

How do children find these basic units? The success of normal children on this task is a highly non-trivial achievement since the physical manifestations of the alleged units are never context-free: In the massively variable signals of natural speech they do not come neatly labelled (Perkell & Klatt 1986). This is all the more surprising when we contemplate that, not only do the units have to be discrete, they also need to be context-free to facilitate recombination effectively.

For the remainder of this chapter we shall try to pinpoint various behavioural skills and factors whose interaction may contribute to the unsupervised formation of large, open-ended and phonemic vocabularies. In so far as they are successful such projects should provide a launching pad for future studies in developmental phonetics.

Some properties of the world's sound systems

Universals and design principles

An examination of the phonetic properties of the world's sound systems offers a useful perspective on the child's learning task. Standard sources like the UPSID database (Maddieson 1984) and the web-based World Atlas of Language Structures (Dryer & Haspelmath 2011) contain phonetic data from a large number of languages. Although these sources give only a sample of all languages spoken around the globe, the selections are made in an effort to make them representative with respect to both geographic dispersion and historical distance.

For the development-oriented phonetician the interest of typological comparisons derives from the possibility that the side-by-side investigation of a really remote and historically unrelated systems will reveal universal patterns and design principles that could provide clues as to the nature of the learning task and how children go about accomplishing that task.

The taxonomy of phonetic properties

Ladefoged & Maddieson's 1996 publication is so far no doubt the most comprehensive account of the articulatory and acoustic properties of the world's vowels and consonants. It summarizes lab research and field work on phonetic contrasts in over 400 languages. In keeping with the taxonomic tradition in phonetics, L & M describe the consonants in terms of their place and manner of articulation. To specify 'place of articulation', they distinguish 17 categories on the basis of two criteria (i) the region of the target place (from labial, dental, … to glottal) and (ii) the articulatory structure that is positioned (their Table 2.1). For example, for targets at the 'dental', 'alveolar' or 'post-alveolar' regions, it is possible to use the tip or the blade as the articulator. As a result their system technically allows for two alveolar 'places': one 'apical alveolar' (involving the tongue tip) and another one 'laminal alveolar' (involving the tongue blade).

Manners are divided up into stops, nasals, fricatives, laterals, 'r-sounds' and clicks. In the case of stops special attention is paid to the state of the glottis (voiced and voiceless), airstream mechanisms (ejectives & implosives) and the coordination of glottal and articulatory activity (pre- & post-aspiration). Nasals present variants linked to the relative timing of oral and velar movements (prenasalized stops, prestopped nasals). Fricatives occur along the entire place dimension and are especially numerous in the dental – postalveolar region since the shape of the tip-blade can be varied in subtle ways. The latter property is also characteristic of laterals and r-sounds. Flaps and taps illustrate the nimble dynamic capabilities of the tongue tip. Clicks show five types: bilabial, dental alveolar, palatal and lateral that can be combined with various glottal actions (voiced, voiceless, breathy, and aspirated).

Vowels are classified in terms 'tongue height', front-back position and lip positions. Additional dimensions include nasalization, advanced tongue root, pharyngealization, frication and retroflection. Phonation types include modal voice, voiceless, creaky voice, stiff voice, breathy voice and slack voice.

Many articulations analyzed by L & M involve a single articulatory movement combined with glottal activity and control of velar opening. In addition somewhat more complex configurations can occur: double closures (labial-dorsals) and secondary modifications (e g labialization, velarization, palatalization, pharyngealization).

On the basis of surveys such as the L & M project and the UPSID database it is possible to make a list of the segment types so far documented. For instance, we can ask: How many different types of /p/? The answer given by UPSID (1984) is: 1. voiceless; 2. long voiceless; 3. palatalized voiceless; 4. voiceless aspirated; 5. long voiceless aspirated; 6. labialized velarized voiceless aspirated; 7. palatalized voiceless aspirated; 8. voiceless pre-aspirated; 9. voiceless with breathy release; and 10. laryngealized voiceless. Moving on to a search through the entire UPSID database we find 556 different consonant segment types and 210 different vowel segment types.

What do these lists tell us? Do they represent hotspots in phonetic space from which languages draw their subsets of contrast?

Systemic constraints

One account of phonetic hotspots is given by the Quantal Theory of Speech (Stevens 1989). It was developed from noting that certain regions in acoustic space are more stable and insensitive to articulatory imprecision than others and then suggesting that languages seek these regions out in the interest of making the acoustic correlates of phonetic units, if not invariant, so at least less variable.

Another way of thinking about finite phonetic inventories starts from assuming that speech perception is not committed to signal invariance. It assumes that correct recognition and comprehension of a spoken word depends on both signal contents and listener knowledge and that the speaker adapts in the short-term to the listener's changing needs for signal information. The acoustic signal produced by such a talker would show variations in the explicitness of the physical cues. Pronunciation would sometimes be clear and rich, sometimes casual and reduced. Therefore the invariance of a given phonetic form would not be expected to be found in the physical signal but only at the level of interaction between signal and stored knowledge that is at the level of comprehension. This reasoning is known as the H&H (hyper&hypo) argument (Lindblom 1990)

According to such a perspective, there are no hotspots. Phonetic values are drawn from a continuous space of possible vocalizations (Catford's anthropophonic space) and unwittingly shaped by speakers, listeners and learners and a tacit demand for robust and reliable communication. Selections are constrained by those factors but are in principle made without limit rather than from a finite set of discrete a priori resources.

One example of such a performance-based criterion is 'contrast', a central concept in phonological analysis. To signal different meanings, words need to be perceptually distinct. Distinctiveness is not an absolute phonetic property. It is a systemic notion. 'Distinctive' means being different from all other elements in a system. The following simplified model of how speech is perceived helps us clarify that point.

When a given word is spoken, it is processed by the listener's auditory system and then interacts with a set of reference patterns stored in memory. The interaction involves the activation of all the items of this mental lexicon in a direct and parallel manner. The degree of activation depends on the similarity of input and reference pattern. The spoken word "resonates with" what is in storage (Shepard 1984). The resonance model suggests that it is the biggest response that determines the identity of the percept.

Two points: Languages are well advised to keep the phonetic shapes of syllables and words distinct because that facilitates the recognition process. Also there is the possibility of listener "expectations" controlling, in a top-down fashion, the sensitivity threshold of the reference elements thereby influencing the processing and thus the distinctiveness of the signal.

The signal is not solely responsible for successful lexical access.

Computer simulations of vowels systems (Liljencrants & Lindblom 1972) lend support to claiming that "*.. a vowel dispersion theory correctly captures a principle governing the distribution of vowels in* *natural languages*" (Disner 1984). The findings suggest that [i e a o u] - the world's most common vowel inventory (Maddieson 1984) – is favoured, not because they individually score high in stability and have certain absolute phonetic values, but because they are sufficiently dispersed in the available space. As a system, they achieve workable robustness in the face of signal degradation and low signal-to-noise conditions.

Contrast is a systemic notion which means that the distinctiveness of a segment, syllable or word depends on how its physical attributes contrast with other competing elements. The cohort constituting the competition is not fixed but varies from moment to moment. The typological evidence strongly suggests that perceptual contrast plays a significant role in the design of both vowel and consonant systems.

The Size Principle

Phonetic systems also bear the marks of productions demands. Ease of articulation is an intuitively appealing explanatory principle but it still lacks a rigorous definition. Nevertheless, there is some indirect evidence that sound systems are shaped by it.

One approach to quantifying this notion was proposed by Lindblom & Maddieson (1988) who performed a sorting of the UPSID 1984 vowels and consonants into three categories: Basic, Elaborated and Complex.

The method was to compare all segments with the respect to multi-tasking - in other words, and to make estimates of the number of articulatory tasks required by any given speech sound. Segments with 'secondary articulations' and 'double articulations' were classified as Elaborated along with those with 'derived' properties: implosives (enhanced voicing), ejectives (reinforcement of release), affricates (extended frication in stop releases), voicing (non-spontaneous) in fricatives and segments involving extreme displacements (retroflexion, subapicals, pharyngeals) and modified default timing patterns (prenasalization, aspiration, prestopping).

Segments having multiple (two or more) extra articulations were labelled Complex. When all single and multiple processes had been identified a small set of Basic segments remained: [p t k ʔ b d g f s h m n ŋ l r w j], all segments involving movement of a single articulation.

When the distribution of Basic, Elaborated and Complex UPSID segments is plotted as a function of inventory size, a systematic pattern emerges: Small systems have mainly Basic articulations. Moderately large inventories have Basic and Elaborated segments. The biggest systems have all three types. Increasing system size entails greater articulatory complexity.

The probabilities of randomly selecting [iᶻ ẽ a̰ o̤ uˤ], or [i e a o u], from the 200 vowel UPSID segment list are equally small, vanishingly so. The Size Principle states that phonetic inventories are products of a tug of war between perceptual contrast and pronounceability. The first of the above systems uses secondary articulations (apicalization, nasalization, breathy voice, laryngealization, and pharyngealization) and could arguably make all elaborated pairs more contrastive than the corresponding basic ones. However, the typological data show beyond all doubt that [i e a o u] is the most frequent system. It appears justified to interpret this to mean that the perceptual advantage achieved by adding the secondary articulations is not sufficient to motivate their extra articulatory 'cost'.

That is the essence of the balance described by the Size Principle which is honoured by both vowel and consonant systems.

'Ease of articulation'

Speech uses a bio-mechanical system with numerous degrees of freedom. Given the rich articulatory and acoustic possibilities there are many dimensions that could be invoked to enhance phonetic contrasts. The typological data indicate that those opportunities are balanced by articulatory factors and left largely unexplored (cf comments on [i e a o u) above). How do speakers go about applying an 'ease of articulation' criterion? Two brief comments.

The 'minimum variance theory' of movement was presented to account for arm and eye movements (Harris & Wolpert 1998). It is based on finding that neural signals tend to be corrupted by noise whose variance increases with the amplitude of the signal (Jones et al (2002). In the presence of such signal-dependent noise, the motor system was observed to adopt a precision criterion that minimizes the variance of the error in reaching the movement endpoint (Hamilton & Wolpert 2002).

Of relevance to the topic of ease of articulation is the fact that high precision implies low-amplitude signals. That strategy also means lower metabolic energy consumption. In other words, the energy cost of the movement is not explicitly monitored but emerges as a by-product of the control.

Both production and learning stand to gain from physiological energy minimization. In speech development, a 'minimum variance' process will bias the child's search for the phonetic forms of the ambient language in the direction of articulations of low complexity. By staying 'easy to produce' a sound system would supplement the child's attempt to imitate what she hears and facilitate the spontaneous production of ambient patterns (cf canonical babbling). An approach of '*Easy-way-sounds-OK*' would contribute to boot-strapping the phonetic development.

On the origin of phonetic units

The syllable

The patterning of the world's speech sounds cannot be investigated without reference to the syllable. Where does this unit come from? The answer given below is inspired by MacNeilage's account of the beginnings of speech (2008): Syllables have their roots in mechanisms developed a long time before human language appeared.

During the past century, attempts were made to specify vowels in terms of their jaw angle, or "Kieferwinkel" (Menzerath & Lacerda 1933). However, classical phoneticians soon lost interest in this parameter when it was pointed out that it is perfectly possible to produce normal sounding vowels with a pencil between the teeth. Since the talker's acoustic output is directly linked to how the tongue shapes the cavities of the vocal tract, a certain lack of interest in the jaw still persists in contemporary phonetics and phonology.

However, the jaw is on its way back. One reason is that, although the tongue clearly shapes the vocal tract, it does so in synergy with the jaw. Jaw movement is a normal attribute of natural speech and its actions have significant articulatory and acoustic consequences (Maeda 1990). So fixed-jaw experiments do not make the jaw irrelevant. Rather it helps highlight the compensatory abilities of the speech motor system.

In fact it turns out that the mandible is central to an insightful description of speech processes (MacNeilage 2008). To reinforce that point we will briefly review two cross-linguistic topics on which jaw movement throws explanatory light: (i) the preferred phonotactic patterns across languages; (ii) the origin of coarticulation.

Phonotactic patterns

Phonotactics is the description of the consonant sequences that are permissible before and after the syllable nucleus (the vowel). The sequencing in clusters reveals a tendency for consonants to order themselves according to their degree of vowel adherence - a quantitative and distributionally based number proposed by Sigurd (1965).

This measure (VAM) is illustrated in Figure 2 with data on initial and final 2-member consonant clusters from Sigurd (1965:49, 74). It is defined as the difference between the number of clusters that have a given consonant in the vowel adjacent position and the number of clusters that have that segment in the non-adjacent position. A positive or negative VAM indicates correspondingly strong or weak vowel adherence. VAM is calculated in the same way for initial and final clusters.

Figure 2 shows VAM values for initial and final 2-member clusters plotted against each other. It can be seen that the measure is high for [r] and [l] and intermediate for nasals, sonorants and stops and lowest for [s].

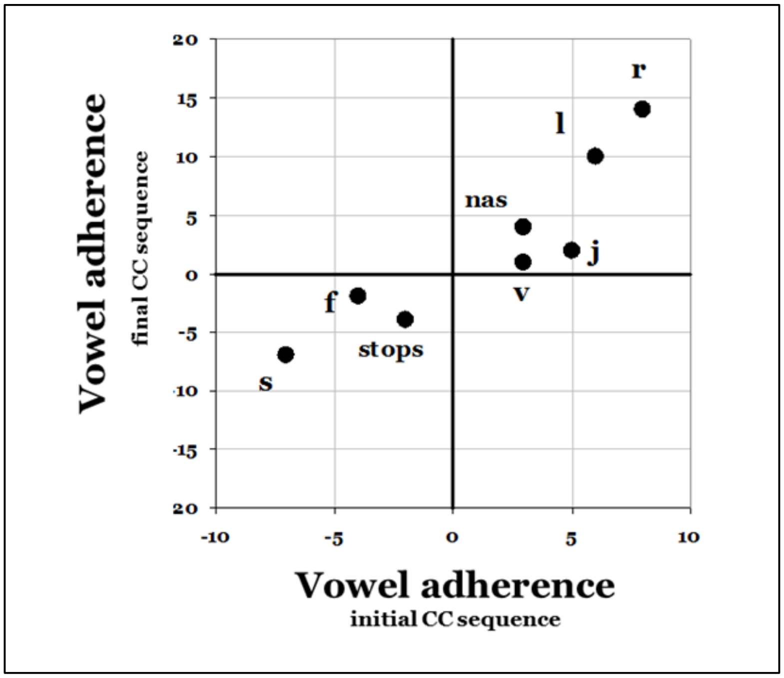


Figure 2 Comparison of vowel adherence values for Swedish CCV and VCC clusters. Abscissa: Initial clusters. Ordinate: Final clusters. Source: Sigurd (1965).

The values for initial and final clusters are similar. This correlation implies that vowel adherence is symmetrical and thus gives rise to mirror image effects; For instance, [skr] occurs initially but not finally, whereas the reverse sequence [rks] is only found finally but not initially. A great many languages conform to the vowel adherence effect.

The following experimental results (Lindblom 1983) give an idea of its phonetic basis.

Subjects were asked to produce symmetrical V1CV2 sequences in which the Swedish fricatives, stops, nasals and liquids occurred in the C slot and a range of vowels with different degrees of opening were used for the V. The movement of the jaw was tracked using the Stockholm University Movetrack system (Branderud 1985). The recorded traces were examined with respect to the position of the jaw in the consonant as well as in the surrounding vowels. As expected the curves all conformed to an open-close-open pattern. The closed positions for the consonants were influenced by the surrounding vowels, most strongly so, between open vowels. The comments below concentrate on the findings for the [’ɑː_a] context which are summarized in Figure 3.

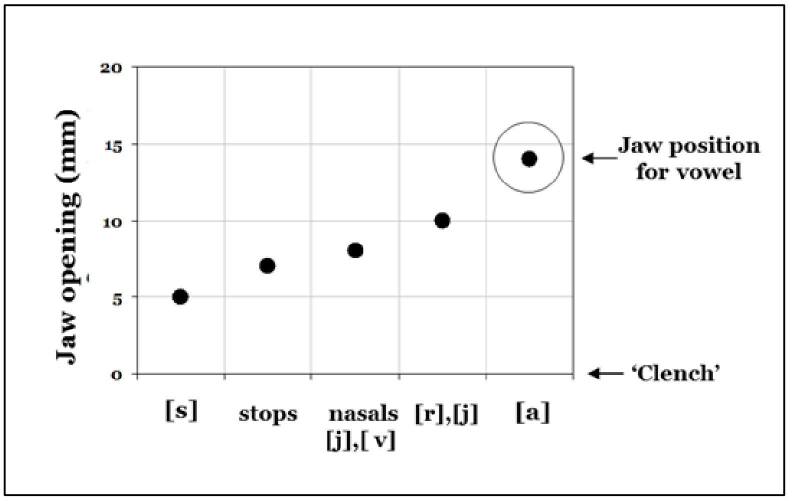


Figure 3 Average jaw opening (in mm relative to clench) for phonetic segments arranged in order of increasing vowel adherence.

As can be seen the tightest closure is observed for [s]. The most open consonant segments are [r] and [l]. Arranging the measurements in increasing order of jaw opening we obtain [s], stops, nasals, sonorants, [r] and [l] – a series that matches the VAM scores perfectly. In these clusters consonants are ordered according to their compatibility with the vowel.

Turbulent noise production in [s] requires air hitting a perpendicular surface with great velocity. Stevens (1998) describes how that happens: "... *The constriction is adjusted so that the airstream emerging from the constriction impinges on the lower incisors*." Conclusion: [s] requires a high jaw.

According to articulatory modeling studies (Lindblom & Sundberg 1971), increasing the jaw opening while keeping the tongue shape fixed will result in a backward movement of the tongue body. Thereby the size of the pharynx cavity is reduced which facilitates making the low and posterior constriction of [a]. Accordingly the natural way to say [a] is with a lowered jaw and to let the tongue and the jaw work in synergy. Conclusion: [a] is an open vowel.

These results suggest that the phonetic basis of vowel adherence is compliance with jaw openness.

Coarticulation in deep time

Further evidence for the jaw's explanatory role comes from segment durations in clusters. A general trend is that the larger the cluster the more the cluster segments are shortened (Haggard 1973). The effect is stylized in Figure 4. The notation is borrowed from music to represent the timing in three Swedish words: [so:r], [spo:r] and [spro:k] (sows, track, language).

Let us compare the three initial [s] segments. In [so:r], its duration is represented by the quarter note. In [spo:r] and [spro:k] it shortens to one half (quaver note) and one third (triplet) of that value respectively. A single jaw trace is drawn to suggest that the jaw is the 'pace maker' that provides the 'beat' common to the three words. Admittedly this is a stylization of the facts but one of sufficient realism for the point we are making.

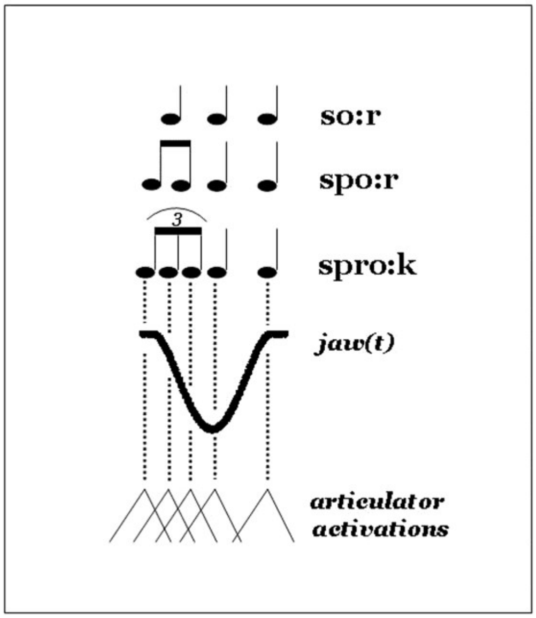


Figure 4 The role of the jaw/ syllable in coarticulation. In musical notation a stylized representation of the timing of the segments in three Swedish words. If a jaw trace common to the three words is assumed the initial consonants are forced to adjust their durations. Vertical lines connect the notes/segments of [spro:k] with the rising and falling lines on the bottom line. They represent the articulatory anticipation/de-activation for [s] [p] [r] etc. The syllable/jaw sets the tempo. Articulatory actions show overlap in time.

The jaw is the 'rhythm section'. The notes (one per segment) indicate the timing of the various articulatory movements. They are not instantaneous. Each one requires periods of anticipation and de-activation. That is indicated at the bottom of the diagram by the saw-tooth shapes of rising and falling lines. On-glides and off-glides take longer than the interval available between the notes. Consequently, the articulatory actions overlap in time which means that there is 'coarticulation'.

Coarticulation arises because of the priority of the basic close-open-close rhythm generated by the jaw. To correctly time the articulatory actions for the initial consonants, the on-glides and off-glides segments must be fitted into the intervals allotted by the jaw-based score. This is done by having the actions overlap, a solution that is facilitated by the use of alternate articulators such as the tongue blade for [s], lips for [p] and the tongue tip for [r].

The old beginnings of the syllable

The open-close alternation of the jaw certainly helps us understand better why there are phenomena such as vowel adherence and coarticulation. However, observations of swallowing and mastication indicate that the jaw is capable of making highly complex motions in precise coordination with tongue activity (Luschei & Goldberg 2011). The oscillation seen in speech represents only a small fraction of its full capacity. Why is its phonetic repertoire so limited and more or less confined to a single dimension?

It is often said that humans lack 'speech organs' proper. Speech is an 'overlaid function' shaped by chewing and swallowing (Kent 2004). Evolution is known to tinker with available resources and to prefer building on existing capacities. Such observations point us in the direction of those capacities for a better understanding of the jaw's speech movements.

The motor systems that became the "vocal" tract were not a tabula rasa but a sophisticated set of tools serving various functions, notably breathing and chewing. It does not appear far-fetched to assume that early speech was significantly moulded by the motor mechanisms already in place (Lund & Kolta 2006).

Breathing, chewing and locomotion exemplify a large class of movements that are cyclic and continuous. Many studies have shown that such behaviours involve central pattern generators (CPG:s) which produce rhythmically alternating motions (Grillner 2006). It appears plausible that the CPG system for chewing came to be used in speech, especially because chewing and speech require the coordination of very similar muscle sets. In speech, precision rather than power is called for. To meet that goal the network would have to be driven at a low-amplitude excitation level (Harris & Wolpert 1998) which - we might speculate - would result in an elementary open-close oscillation near the mandibular system's resonance frequency and which would satisfactorily meet precision demands and communicative needs.

Summarizing this section we make two points: (i) Evidence was presented for giving the open-close alternation of the jaw a key role in explaining preferred cross-linguistic phonotactic patterns and identifying the source of the spatial and temporal overlap of articulators known as coarticulation. Although reinforcing belief in the importance of the jaw, this result still left the question unanswered of where the open-close alternation comes from. For a reply we adopted a phylogenetic perspective concluding that the alternation is a natural product of using the motor mechanisms on which speech is an overlaid function - a pianissimo gently performed on the CPG keyboard of mastication. This natural motion captures the phonetic essence of the 'syllable'.

The discreteness of phonetic segments

Clues from the IPA.

The practice of describing speech sounds in terms of static articulatory attributes can be traced far back in phonetic history. The International Phonetic Association provides the world with the notational standard for the phonetic description of all languages. Its alphabet, revised in 2005, is presented in charts containing the current set of phonetic symbols and diacritics. Consonants are specified with respect to manner and place and voicing, vowels in terms of front-back, open-close and rounded-unrounded. For phonation types and secondary articulations diacritics are used.

Do the descriptive labels given to the dimensions of the IPA (i.e., place, manner, source etc. categories) reveal anything about how the different segments are produced by the motor system? Most of the classificatory terms refer to static states of vocal tract (such as 'voiced', 'voiceless', 'labial', 'dental', ' uvular', 'stop', 'nasal', 'fricative' …) implying a target-driven rather than gesturally based organization of speech production. For example, the slots along the place dimension could be interpreted as static positional goals for the motor system and the motions from and to those regions could be derived as secondary consequences. The patterns representing the acoustic targets of manner and source features could be similarly defined in terms of discrete steady-state vocal tract configurations.

Clues from experimental phonetics

The term 'target' has been widely used with reference to articulatory positions or acoustic patterns that a talker aims at but does not always reach (Stevens & House 1963). It continues to produce the most parsimonious descriptions of complex articulatory and acoustic data.

An example is the recent set of measurements of the articulatory and acoustic properties of bV, dV and gV syllables reported by Lindblom & Sussman (2011). The goal of the research was to predict the coarticulatory patterning of stop-vowel formant transitions by means of a computational articulatory model. The observed effects were used as a window on the underlying control processes. Good quantitative accuracy was obtained by using a single target configuration for each stop and each vowel and by generating the transition as a smooth trajectory with fixed shape and time constant and scaled to interpolate between the consonant and vowel targets. These results extend and reinforce Öhman's classical target-based account (Öhman 1967).

The developmental implication of these two studies is that they reveal the end state of learning to produce and coarticulate stops correctly: For each place of articulation the child needs to find the unique context-independent target – the representation that sounds right and generates the coarticulation patterns of the input. The modelling demonstrated that, once the onset and endpoint were known, any given CV transition was fully specified. This further implies that, once the child gets the correct spatial targets set up, the articulatory and acoustic properties of the transition will unfold automatically as a result of the non-linear articulation/acoustics mapping and the general motor mechanisms of trajectory formation (Flash & Hogan 1984, Shadmehr & Wise 2005).

Clues from neurobiology

Many non-speech continuous motions produced by CPG networks also allow superposition of discrete positioning. An example is the special case of locomotion known as '*precision walking'*. It requires the visually guided coordination of the rhythmic CPG activity and the signals for the discrete movement generated by neurons in the motor cortex in precise phases of the movements (Grillner 2006, Ijspeert 2008).

A similar analysis is presented for reaching tasks. "The motor cortex and its corticospinal outflow are preferentially engaged when precise positioning of the limb is needed during locomotion and are also involved during reaching and active positioning of the hand near objects of interest" (Georgopoulos & Grillner 1989).

It does not seem far-fetched to fit speech into this framework. Paraphrasing the formulation we just used above, we could describe it as requiring auditorily (rather than visually) guided coordination of the rhythmic CPG activity (read: the syllabic rhythm of the jaw) and the signals for the discrete movement (the phoneme command) generated by neurons in the motor cortex in precise phases of the (jaw) movements.

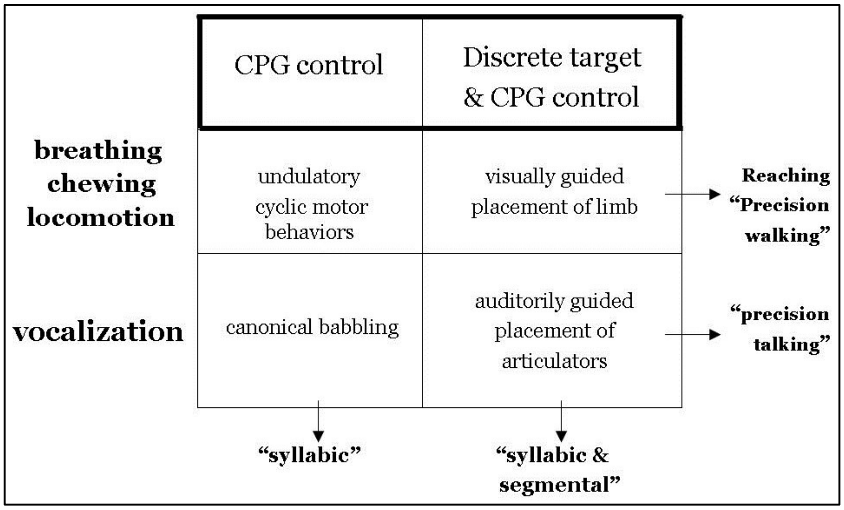


Figure 5 Top row describes non-speech motor mechanisms. Bottom row indicates how those mechanisms are paralleled in speech (from Lindblom & MacNeilage 2011).

Figure 4 highlights the possible parallels between speech and non-speech movements. In the top left cell examples of CPG-based rhythmic movements such as breathing, chewing and locomotion are exemplified. In the right hand cell we find locomotion that involves exact foot placement: '*precision walking'*.

The bottom cells extrapolate to speech. Canonical babbling is here assumed to be driven by CPG networks also serving chewing. Infants do not have sufficient control to produce genuine segments but gradually become able to coordinate the basic jaw movement with articulatory activity. Rather than just letting the jaw close the lips as in canonical babbling, the child may intentionally reinforce that labial occlusion by active muscular action, or may, with time, successfully make it elsewhere using a different articulator. As this placement of articulators is generalized it represents a form of discrete target control - ‘precision talking’ - a close parallel to ‘precision walking’.

We hypothesize positional control as the source of discreteness in phonemic units. It is a general process that is available to the child and would facilitate the launching of the discrete coding of speech patterns accordingly taking the role of an important boot-strapping function.

Vocal imitation and analysis-by synthesis

Children learn to link speech percepts from the ambient language to their own articulatory actions. The possibility of a neural account of vocal imitation is suggested by the discovery of mirror neurons (Gentulucci et al 1988, Rizzolatti et al 1988). These neurons which were first identified in the macaque’s premotor cortex, discharge when the monkey manipulates objects and when it observes other monkeys or humans do the same. Neurons that respond to sound and to communicative or ingestive actions have also been identified (Ferrari et al 2003). In humans the existence of a cortical network with the properties of mirror neurons has been demonstrated (Cattaneo & Rizzolatti 2009, Pulvermüller & Fadiga 2010). Brain stimulation and imaging studies indicate increased activity in speech muscles when subjects listen to speech (Fadiga et al 2002; Watkins et al 2003, Wilson et al 2004).

This sound-to-movement mapping is obviously an important aspect of phonetic learning but there is an issue concerning its resolution. The question is whether its developmental role extends to all the fine details of speech signals. It appears reasonable to look also for other sources aiding children to emulate what they hear. For instance, consider ambient patterns that are simple for the child to produce. They may spuriously appear to be imitations but could in fact be produced spontaneously because of their sheer naturalness - cf the pseudo-syllabic nature of canonical babbling. In other words, if sound patterns keep their articulatory complexity low, chances are that, during his exploration of the phonetic space, the child might accidentally score quite a few hits (‘*Easy way sounds OK’*).

The child’s path to a phonemically organized vocabulary might be a case where such an (unintended) analysis-by-synthesis strategy would come into play and yield significant pay-offs.

Phonetic recombination

Motor tools

The insights we need to make better sense of phonetic recombination in children come once more from research on non-speech motor mechanisms. Parsimoniously, we assume that motorically speech is not special. Therefore what we learn from this domain should apply also to speech.

We highlight two aspects: (i) The organization of motor systems is output-oriented. Motor equivalence (Lashley 1951) exemplifies this type of control; (ii) Movement paths (as in point-to-point reaching) exhibit certain invariant kinematic properties including ‘minimum jerk’ and bell-shaped velocity profiles (Flash & Hogan 1984, Shadmehr & Wise 2005).

The conclusion from the first topic is that spatial goals can be reached from arbitrary onset points provided that onset and endpoint are located within the system’s work space. From the second topic we learn that movement paths can be scaled according to the extent and duration of the task but their shape and time constant tend to be fixed across a large range of conditions.

With respect to speech development the implication is the following: When the child puts these capabilities into play they will enable her to make the appropriate transition between any two speech sounds provided that she has mastered their target representations. In other words, she is ready to "recombine".

Motor equivalence

The classical example of motor equivalence is handwriting (Lashley 1951). Letter shapes tend to stay constant whether one writes on a piece of paper, a blackboard or in wet sand making the strokes with the foot. Such examples show that motor control is output-oriented: There is constancy in the outcome although each time the motor system has to recruit different sets of muscles to get the job done. The motor implementation depends on the context but the goal stays the same.

The task of reaching for a glass in a specific location is carried out by different muscles depending on whether the glass is above, below, in front of or behind the person performing the action. Motor systems evolved to operate in a compensatory mode. They are set up to deal with the unforeseen. Motor equivalence refers to the ability of the system to move an effector (arm, foot, articulator …) from its current location A to an arbitrary position B within the work space.

Kinematic constancies of non-speech (& speech) movement paths

By parsing the continuous flow of speech into strings of targets, we treat it as a sequence of reaching gestures. A lot of research has been done on point-to-point reaching movements (Shadmehr & Wise 2005). A salient fact about such movements is that they are highly replicable from trial to trial and from subject to subject. The trajectory approximates a straight line. Its velocity profile is consistently a unimodal, bell-shaped curve. Those features are retained when position, extent and duration are varied. The big questions are how and why the brain arrives at a unique solution each time with the above characteristics. The arm-hand system has many degrees of freedom and subjects are certainly capable of using them to make various more "unnatural" motions. Accordingly there are strong indications that motor systems arrive at uniqueness by adopting criteria that limit the search space to a single optimal choice (Soechting and Flanders 1998). That is presumably also true for speech gestures.

The classic study by Flash & Hogan (1984) presents a model of point-to-point hand trajectories. They found that the computed movements matched observed movements with high accuracy when they minimized the derived curve's rate of change of acceleration - the so-called 'minimum jerk' criterion. They remarked that "*the predicted trajectory depends only on the initial and final positions of the hand. It is invariant under translations and rotations. Thus the shape of the predicted trajectory does not change with amplitude or duration of the movement, which merely serve to change the scale of the position and time axes, respectively*".

Ready to build an open-ended vocabulary

As her motor development progresses, the learner can use a) positional target control, b) her ability to produce motor equivalent movements and c) her ability to make minimum jerk transitions. These tools and skills will enable her to represent phonetic input as strings of context-free control objects (targets). It will also obviate the necessity to include details on the articulatory properties of the transitions. They will unfold automatically as a result of how motor mechanisms handle trajectory formation.

Assume a child who consistently uses [ba] for 'ball' and [dɛ] for 'daddy'. She often hears the words 'bed' and 'dog' but does not yet attempt to say them. However, since she is familiar with the stop and vowel targets of ball' and daddy', she is in principle ready to recombine them and generate the new forms [bɛ] for 'bed' and [da] for 'dog'.

The key idea of the present account is captured by the phrase "once the targets … have been established". If the target assumption is accepted, the construction of an open-ended lexicon by recombining context-free discrete units becomes less of a mystery. We have come a bit closer to being able to deal with the questions posed in the introduction by proposing that: Positional control paves the way for discrete units; Motor equivalence offers context-independence; Motor equivalence and minimum jerk movement paths produce the appropriate movement between any two contiguous targets.

Final piece of the puzzle: Where re-use comes from

If we test what we have so far by also asking "Where does re-use come from?" we realize that our sketch addresses that issue only in part. It would be fair to say that we now have an idea of how re-use gets *maintained* by each child. However, our account does not get to the ultimate source of the phenomenon. It leaves children's development of re-use to the ambient input. In other words, since all languages exhibit that feature, it is not surprising that it shows up also in the acquired lexicon of each individual. To answer more fully it will be necessary to seek an evolutionary answer.

Although we cannot make direct observations of the early steps taken by our ancestors towards phonetic re-use in vocabularies, it appears likely that they occurred in the first attempts at creating and memorizing new words. According to Donald (1991) a communicative 'mimetic' culture arose during the period of Homo erectus allowing individuals to share mental states and to represent reality in new and expanding ways. Man's growing cognitive capacity created a strong demand for fast and precise ways of communicating. In Donald's estimate spoken language appeared after the period of archaic Homo sapiens in response to such demands. An abundance of meanings had to be paired with their individual phonetic shapes – either by the invention of holistic signals, or by building on the vocal patterns already in use.

Evolution's preference for tinkering with existing resources makes the latter route more probable. New meanings had to sound different. Pre-linguistic vocalizations were subject to same production constraints that we have found in present-day speech and sound systems. When new things had found their names they had to be committed to memory. It is therefore likely that the available vocal patterns already formed a small subset of the 'anthropophonic' space, i.e., man's total sound production capacity (Catford 1982). The playing field was not level with respect to that total capacity. The phonetic search space would have been reduced by tacit preferences of 'easy to say', 'easy to hear' and 'easy to learn'. To come up with phonetic shapes for a large number of meanings, would involve making forced choices and occasionally re-using the discrete elements of the available resources.

After reaching the canonical babbling stage infants become better able to coordinate the basic jaw movement with articulatory activity. There then follows a period when the child explores its capabilities by abandoning the passive, default behaviour involving just the jaw and begins to vary its placement of articulators (variegated babbling). The moment when, stimulated by what she hears around her, the child intentionally reinforces the labial occlusion of the babble by active muscular action, or uses the tongue to make the closure elsewhere, she takes the first step towards discrete positional control - 'precision talking'. As her exploration continues several more targets are born. It is not inconceivable that, at a distant point in time, human vocalizations took a similar path.

In terms of bit rate the information of a holistic vocal pattern is expensive. By definition its overlap with other patterns is minimal and therefore its specification takes a lot of bits. Modifying an existing item is cheaper since the overlapping parts come free of charge. A physiological basis can be found for this economy argument. In biology memory storage is associated with a biochemical 'cost' which derives from the energy metabolism of memory formation. Brains change physically as a result of learning. This change is activity-dependent. Active neurons contain more energy-rich substances. Should a demand for their activation arise (e.g., recall), active neurons have the 'fuel' to respond. Building up that capacity costs metabolic energy.

It is possible to measure the effect of learning on brain tissue. A substance used as an index of metabolic capacity is cytochrome oxidase. More active neurons have greater amounts of this enzyme. Experiments were run on rats trained to associate reward with an auditory stimulus. Histochemical analyses of brain tissue were then performed on experimental and control animals. The experimental group showed significantly increased amounts of cytochrome oxidase in the auditory neostriatum. The memory of the conditioning stimulus had modified the neurons activated by the task (Gonzales-Lima 1992).

The implication of this type of work is that patterns that contain more information and whose specification therefore requires more bits, are energetically more costly, and therefore they take longer to commit to memory. The new phonetic forms that would be especially favoured by the metabolic factors of memory formation would be the ones that conform to a '*minimal incremental storage'* criterion. Such considerations give some substance to the notion of 'easy to learn'.

With vocalizations under positional control in the repertoire, opportunities would have arisen for evolution to apply the re-use method. Once it had been introduced it would have been maintained and reinforced owing to the numerous advantages it offered: Its emergence can be seen as a case of selection-by-consequences.

Conclusion

The goal of this paper was to demonstrate the feasibility of a non-structuralist approach to the basic units of speech. Accordingly we reversed structuralist priorities starting from substance (performance-based factors) in a quest for possible behavioural causal factors underlying sound structure formation. We found that a strong case can be made for the claim that the combinatorial structure of sound patterns has behavioural origins.

References

Anderson S R (1981): "Why phonology isn't "natural"", *Linguistic Inquiry 12*:493-539.

Anderson S R (1985): *Phonology in the twentieth century*, Chicago: Chicago University Press.

Branderud P. (1985). "Movetrack - A movement tracking system". *Perilus IV*, Stockholm University, 20-29.

Catford J C (1982): *Fundamental problems in phonetics*, Indiana University Press: Bloomington.

Cattaneo & Rizzolatti (2009): "The mirror neuron system", *Arch Neurol 66(5)*:557-560.

Chomsky N (1964): "Current trends in linguistic theory" 50-118 in Fodor J A and Katz J J (eds): *The structure of language,* New York:Prentice-Hall.

Disner S (1984): "Insights on vowel spacing", 136-155 in Maddieson I (1984): *Patterns of sound*, Cambridge.

Donald M (1991): *Origins of the modern mind*, Cambridge, MA: Harvard University Press.

Dryer, M S & Haspelmath M (2011): *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, Available online at http://wals.info/

Edlund S (1957): *Studier rörande ordförrådsutvecklingen hos barn i skolåldern* (Studies of vocabulary development in school children), Lund, Sweden, Gleerup.

Flash T & Hogan N (1984): "An organizing principle for a class of voluntary movements", *Journal of Neuroscience 5(7)*: 1688-1703.

Georgopoulos A P & Grillner S (1989): Visuomotor coordination in reaching and locomotion", *Science* 245:1209–1210.

Gonzales-Lima F (1992): "Brain imaging of auditory learning functions in rats: Studies with fluorodeoxyglucose autoradiography and cytochrome oxidase histochemistry", 39-109 in Gonzales-Lima F, Finkenstädt T & Sheich H (eds): *Advances in metabolic mapping techniques for brain imaging of behavioral and learning functions*, NATO ASI Series D:68, Dordrecht: Kluwer.

Grillner S (2006): "Biological pattern generation: The cellular and computational logic of networks in motion", *Neuron 52*:751–766.

Haggard M (1973): "Abbreviation of consonants in English pre- and post-vocalic clusters", *J of Phonetics* 1(1):9-24.

Hamilton A F & Wolpert D M (2002): "Controlling the statistics of action: Obstacle avoidance", *Neurophysiol 87*:2434–2440.

Harris C M & Wolpert D M (1998): "Signal-dependent noise determines motor planning", *Nature* 394:780-784.

International Phonetic Association (2005): http://www.langsci.ucl.ac.uk/ipa/

Ijspeert A J (2008): Central pattern generators for locomotion control in animals and robots: a review, *Neural Networks 21/4*:642-653.

Jespersen O (1926): *Lehrbuch der Phonetik*, Leipzig:Teubner.

Jones K E, Hamilton A F & Wolpert D M (2002): "Sources of signal-dependent noise during isometric force production", *Neurophysiol 88*:1533–1544.

Kent R D (2004): "Development, pathology and remediation of speech", in Slifka J, Manuel S & Matthies M (eds) (2004): *From sound to sense: 50+ years of discoveries in speech communication*, conference proceedings (CD-ROM format), Research Laboratory of Electronics, MIT, Cambridge, Mass.

Kroos C, Geumann A & Hoole P (1999): "Tongue–jaw trade-offs and naturally occurring perturbation", *J Acoust Soc Am 105(2)*:1355-1355.

Lashley K (1951): "The problem of serial order in behavior", 112-136 in Jeffress L A (ed): *Cerebral mechanisms in behavior*, Wiley:New York.

Lindblom B (1990): "Explaining phonetic variation: A sketch of the H&H theory", 403-439 in Hardcastle W J & Marchal A (eds): *Speech Production and Speech Modeling*, Dordrecht:Kluwer.

Lindblom B & MacNeilage P F (2011): "Coarticulation: A universal phonetic phenomenon with roots in deep time", *FONETIK 2011*, 41-44 in TMH - QPSR Vol. 51, KTH Stockholm, http://www.speech.kth.se/prod/publications/files/3588.pdf

Lindblom B & Sussman H M (2012): "Dissecting coarticulation: How locus equations happen", *J of Phonetics 40(1):*1–19.

Lund J P & Kolta A (2006): "Brainstem circuits that control mastication: Do they have anything to say during speech?", *J Communication Disorders 39(5)*:381–390.

Luschei E S & Goldberg L J (2011): "Neural mechanisms of mandibular control: Mastication and voluntary biting", *Comprehensive Physiology* 1237–1274.

Maddieson I (1984): *Patterns of sound*, Cambridge.

Maeda S (1990): "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model", 131-149 in Hardcastle W J & Marchal A (eds): *Speech Production and Speech Modeling*, Dordrecht:Kluwer.

Menzerath P & Lacerda A (1933): Koartikulation, Steuerung und Lautabgrenzung, Berlin.

Miller G A (1977): Spontaneous apprentices: Children and language, New York: Seabury.

Miller G A (1991): *The science of words*, New York: Freeman.

Öhman S E G (1967): "Numerical model of coarticulation", *J Acoust Soc Am 41*:310-320.

Perkell J S & Klatt D H (1986): *Invariance and variability in speech processes*, Hillsdale, NJ: Erlbaum.

Shadmehr R & Wise S P (2005): *The computational neurobiology of reaching and pointing*, The MIT Press: Cambridge MA.

Shepard R N (1984): "Resonant kinematics of perceiving, imaging, thinking and dreaming", *Psychological Review 91(4)*: 417-447.

Sigurd B (1965): *Phonotactic structures in Swedish*, Lund, Sweden, Gleerup.

Stevens K N (1989): "On the quantal nature of speech," *J Phonetics 17*:3-46.

Stevens K N (1998): *Acoustic phon*etics, Cambridge MA: M.I.T. Press.

Stevens K N & House A S (1963): "Perturbation of vowel articulations by consonantal context. An acoustical study", *J Speech Hearing Res* 6:111-128.

Teleman U (2012): "Ordförråd", *Nationalencyklopedin*.

Particulate Speech:

The Emergence of the Phoneme from Syllable Frame Structures

Peter F. MacNeilage
University of Texas at Austin
macneilage@psy.utexas.edu

Abstract

William Abler (1989) has postulated the "particulate principle of self-diversifying systems … which holds that when systems form structures having an infinite range of properties, such systems must be based on particles …". According to Abler, natural systems exhibiting this principle "include chemical interaction, biological inheritance and human language". The present focus will be on the linguistic particles singled out by Abler—the phonemes of language. According to Abler's physics-based conception, the three systems "derive their properties by common inheritance" from "some property of material nature". However, in the present view physics and biology differ in that the source of ultimate causes in biology is the theory of evolution by natural selection, and phonemes must be explained in terms of this theory, not in terms of the laws of physics. As to the origin of the three instances of particulation, the particles of chemical interaction, and biological inheritance—elements and nucleic acids respectively—were synthesized from pre-existing subcomponents, while phonemes may have emerged from a pre-existing syllable frame structure, according to the author's "Frame/Content" theory of evolution of speech production, reviewed here. Two subphonemic particles of speech have also been proposed—the "distinctive feature" of structural linguistics and the "gesture" of proponents of dynamical systems theory. However, these particles do not exist in nature. In conclusion, syllable frame structures can be considered to provide a unique pre-systematic basis for both the evolution and the development of phonemic systems of languages.

Keywords: speech; syllable; phoneme; gesture; distinctive feature; evolution; acquisition; comparative neurobiology, essentialism; natural selection

Introduction

William Abler has postulated the *"particulate principle of self-diversifying systems … which holds that when systems form structures having an infinite range of properties, such systems must be based on particles …."* (Abler, 1989, p. 1). According to Abler, *"These systems include (1) the chemical elements, whose underlying particulate units are atoms, (2) biological inheritance, whose particulate units are genes, and (3) human language, whose underlying particulate units are phonemes and morphemes" (p. 1)*. Abler designates the three systems as *"Humboldt Systems"* in that they *"share the further property of conforming to Humboldt's (1836) characterization of language, i.e., all systems 'make infinite use of finite media' (p. 70) whose synthesis creates something that is not present per se in any of the associated constituents' (p. 67)" (p. 1).*

The present author concurs with Abler's view that the relevant particle in linguistics is the phoneme. A phoneme is a sound—a consonant or vowel—that differentiates one word from another. For example, the words "ten" and "den are different by virtue of their initial consonants, denoted in phonemic transcription as /t/ and /d/. Self-diversification in this context refers to the way in which segments are concatenated in many ways to form the meaning units, or morphemes of words. According to the author's Frame/Content Theory (MacNeilage, 1998; 2008) phonemes (collectively content) emerge, in both evolution and development from a superordinate syllable frame structure. This theory, and the evidence for it will be presented, before considering two other putative speech units, the "distinctive feature" and the "gesture", and before considering Abler's proposed explanation for the existence of the particulate principle.

The Frame/Content Theory

Phylogeny

Speech can be characterized as a voice source (described as phonation or voicing) modulated by a series of mouth closing/opening alternations. An alternation cycle is a syllable, and the closing and opening phases are consonants and vowels, respectively. In an important paper entitled "The Problem of Serial Order in Behavior" Lashley (1951) noted that evidence regarding the serial organization of speech comes from serial-ordering errors in normal speakers because these errors not only reveal the units of speech—phonemes, consisting of consonants and vowels—but also the constraints on how they are ordered. In these errors (e.g., spoonerisms such as "catnip" → "natkip" and "ad hoc" → "odd hack"), misplaced consonant and vowel forms (segments) end up in the same position in syllable structure that they originated in. The widespread nature of this syllable structure constraint on segmental positioning led Levelt, the most prominent figure in the psycholinguistics of language production, to conclude that "Probably the most fundamental insight from speech error research is that a word's skeleton or frame and its segmental content are independently generated" (1992, p. 10; italics mine). The Frame/Content theory is a theory about the origin of this form of organization, in both phylogeny and ontogeny.

One might ask, what is so remarkable about this form of organization. Well, for one thing, it doesn't happen in typing, a linguistic serial-ordering operation that we invented. Here, we are just as likely to reverse consonant and vowel letters as we are to reverse pairs of consonant letters, or pairs of vowel letters (MacNeilage, 1985). So the fact that, for speech, nature has chosen something different from a simple linear ordering of units (as in keystrokes) that occasionally suffers from little misalignments of these units seems to be an important clue as to how we came to be able to do speech.

According to the Frame/Content theory, the syllable frame evolved from cycles of mouth close-open alternation (closed for consonants, open for vowels) controlled by the mandible. A cyclicity of this general form evolved in mammals for acts of ingestion (chewing, sucking, and licking) about 200 million years ago, and we probably exapted (borrowed) it for speech. As the two movements in the cycle are antagonistic—they go in opposite directions in successive phases of the cycle—and as the two major-class segmental categories (consonants and vowels) probably evolved from these two opposite movements, there was never an opportunity for consonants and vowels to get mixed up with each other in the control of the cycle, hence the syllable structure constraint on speech errors. Following the evolution of ingestive cyclicities, there may have been an intermediate stage of use of this cycle for visuofacial communicative cyclicities (e.g., lipsmacks, tonguesmacks, teeth chatters). These communicative sequences of close/open alternations are common in modern monkeys (Redican, 1975). In the step from smacks to speech, the cycles of mandibular oscillation were paired with vocal-fold vibration (voicing, phonation) to form protosyllable sequences such as "bababa." The suggestion that protosyllables like this may have been the original phonological forms of speech is endorsed in the recent influential monograph by Jackendoff (2002).

How was this putative sequence of events chosen by natural selection. The advent of internal temperature regulation in the transition from reptiles to mammals required higher rates of food ingestion, and the original mandibular cyclicities for chewing, sucking, and licking can all be regarded as increments in food-processing efficiency selected for that end. Van Hooff (1967) has suggested that smacks may have evolved their communicative status from cyclical ingestive movements elicited in a manual-grooming context. Animals looking forward to finding a food item such as a salt grain in individual instances of grooming might have begun chewing movements in anticipation of such a discovery. Initial selection pressures for the protosyllabic cyclical forms (in effect, smacks with voicing) may have come from vocal grooming, whereby, according to Dunbar (1996), vocalization may have been substituted for manual tactile contact as ancestral hominid group sizes increased enough to make the latter behaviour ineffective as a device for social cohesion. (Both manual and vocal grooming are widely regarded as having beneficial consequences for social organization.) It's also possible that an evolving capacity to learn vocalization—a capacity not present in other living primates—occurred when vocal grooming became important as part of a general-purpose mimetic capability, selected to enhance group solidarity, as suggested by Donald (1999). This capability to recreate the observed actions of others, almost as salient and unique in humans as is speech, is evident in modern human music, dance, opera, ballet, movies, games, sports, etc. However, there has been a marked reluctance to give Donald's claim of a fundamental status for this capability the respect it deserves in the domain of human phylogeny.

Ontogeny

According to the frame/content theory, the process of speech acquisition may throw light on its evolution. A compelling consideration here is how rhythmic infant babbling is from its very beginning, a basic action available to all normal infants. There is no sign of a process whereby infants gradually make their output more rhythmic as there would be if they were learning this capacity, in the way, for example, that one learns to make a smooth, rhythmic golf swing. For example, we have found that 2 successive close-open cycles of babbling, which average 1/3 of a second each, differ from each other, on the average, by only 1/30th of a second (Dolata, Davis and MacNeilage, 2008). The degree of precision in this cyclical action of a 7 month old infant, from its very beginning, even though it is not a life supporting action, unlike the heart beat or respiration, suggests that the underlying function has an extremely fundamental status. Furthermore, as we will shortly see, the biomechanics of this early speech-like output of young modern hominids are extremely simple, suggesting that the biomechanics of the original ancestral speech-like actions must have been quite simple too, and, of course, similar. This, among other considerations points to the conclusion that babbling has evolved as an innate substructure for speech (MacNeilage, in press).

Over a generation of work on speech acquisition with Barbara Davis and others (see, for example, MacNeilage and Davis, 2000) has led us to the conclusion that the initial babbling stage of speech development, in which the same syllable tends to be "reduplicated" (as in "bababa"), can be regarded as a "Frame" stage, with little in the way of internal "Content." We have found, and others have confirmed, that the production of both babbling and early words is mainly a matter of oscillation of the mandible, giving rise to an alternation between a closed- and open-mouth configuration. In a number of studies we have shown that sequential adjustments of tongue position, which is the main determinant of the different consonants and vowels of a language, tend not to occur in early babbling and speech episodes (See Davis and MacNeilage, 2002 and MacNeilage and Davis, 2002 for summaries).

This is a sufficiently central property of speech acquisition to deserve some detailed illustration. For example, if the tongue is in the front of the mouth when an infant begins a babbling episode or a word, it tends to stay there throughout. Consequently, tongue-front consonants such as [t], [d] and [n] tend to accompanied by front vowels in consonant-vowel (CV) syllables. Examples of words with these form are "Ted," "dead," and "Ned." A similar relation is found between tongue-back consonants ([k] and [g]) and back vowels, as in words such as "cook" and "gook." Perhaps most interestingly, the vowels produced following lip consonants (e.g., [p], [b], [m]), such as the first vowel in "papa" and the vowel in "bubble," and the so-called neutral vowel in the first word of "the box" are called "central vowels" and are probably made with the tongue in its resting position in the center of the mouth. In all these cases, the consonant-vowel sequence, even when it's repeated a number of times, does not involve active tongue movement at all. We have called this overall phenomenon "Frame Dominance" (Davis and MacNeilage, 1995, p. 1199).

Thus, apart from the cyclical frame, the operation of the speech apparatus in babbling and early words is characterized primarily by biomechanical inertia. The tongue is apparently not called upon to change its position during a vocal episode. Davis and I reasoned that if these inertial patterns were also favoured in modern languages, this would constitute evidence for their fundamental status in the evolution of speech. In our study of 10 languages (MacNeilage et al, 2000) and in a study of 14 languages by Rousset (2003), the three basic patterns described above were present in about 70% of instances, compared with an expected base rate of 33%

There is one other basic relation between babbling and speech patterns of language that suggests the former is fundamental to the latter. It is that the consonant-vowel (CV) form favoured in these infant vocalizations is, with a single exception, (Breen and Pensalfini, 1999) the only universal syllable type in the world's languages.

If indeed the first stage in the evolution of speech was a frame stage, how then did we eventually evolve the capacity to program individual segments into syllable frame structures using a frame/content mode, thus achieving "segmental independence (MacNeilage and Davis 1990b)? A clue as to how this might have begun to happen comes from the course of events in speech acquisition. There is a strong trend across language environments for the first systematic step toward intersyllabic variegation to involve a preference for starting a word with a lip consonant and following it after the vowel with a tongue-front consonant—for example, in the word [bado] for "bottle." In studies of adult languages, my colleagues and I, (MacNeilage et al., 2000) as well as Rousset (2003), find that, as in the case of the tongue-inertia tendencies, this tendency is also rife in languages. Davis and I have suggested that this is a result of self-organization (MacNeilage and Davis, 2000). It may be a case of beginning simply with what we call a "Pure Frame," (a lip consonant-central vowel sequence produced only by the mandible) because, in lay terms, it is easier. It doesn't also involve beginning with a non-resting tongue position. But once having begun, the infant can then take the additional step of making the tongue movement needed to get the tongue-front consonant. This conjecture is supported by evidence from neurophysiology that starting to move is a special problem for the brain, because it is addressed with dedicated circuitry (e.g. Gazzaniga and Heatherton, 2003).

At a more general level, the emergence of the segments comprising content from syllable frames during the course of early development can be viewed in the following way. Consider first the case of pure frames in which no independent tongue action accompanies the mandibular cycle for the CV syllable. Resulting in a labial (lip) consonant and a central tongue position for the vowel. What is necessary, in order to produce this kind of consonant and vowel in adult speech along with other segments, at the rate of about 15 segments per second is that we develop the capacity to actively control lip closure for the consonant and tongue control for the assumption of the required central tongue position. Electromyographic studies of muscle action potentials during adult speech show the presence of this active command structure (e.g. MacNeilage and DeClerk, 1969). Obviously the commands to the muscles must be restricted to the time windows allowed for the movement associated with the labial consonant and the movement associate with the central vowel at this speaking rate. Once this active control of individual labial consonants and central vowels has been achieved, these segments can be produced in contexts other than the one in which they co-occur with each other, as passive resultants of mandibular oscillation.

Consider now the development of control of individual content segments from the beginning point of the favoured frames that involve the tongue (coronal-front and dorsal-back CV forms). In this case, when the lingual consonants and their preferentially associated vowels are not called for together in target words, it is necessary to learn to restrict the tongue positions associated with the individual segments to single consonants (coronal or dorsal) and single vowels (front or back).

In summary, at the level of peripheral control, the ability to program individual segments in the case of pure frames requires generation of segment-specific movement control capacities where none existed before. In contrast differentiation of segments from lingual frames requires the restriction of the tongue positioning movement to the time window of a single segment, be it the consonant or the vowel. In the particular case in which the LC sequence is in the word being attempted, this ideally means restricting the tongue movement to the final consonant if the vowel is not a front vowel, though one might expect, at early stages that there might be some tendency for the vowel preceding the final consonant to share its front place of articulation because it might take some time to reliably make the movement specific to the coronal consonant. (See Lindblom and MacNeilage, 2011, for further discussion.)

Davis and I have not pursued the specifics of what we have called the “Frames, then Content” progression (MacNeilage and Davis, 1990a) of the infant/child in speech acquisition beyond the identification of the labial-coronal effect. The obvious question to ask is whether, as in the case of adults, serial ordering errors of infants throw light on this progression. Fortunately, we have available a large scale study of speech errors during speech acquisition. Jaeger (2005) has analyzed 1,383 speech errors made by 3 infants/children within the age range of 1-5 years. In a summary of her findings she states that “An extremely valuable heuristic in which to frame this discussion is the ‘Frames, then Content’ developmental progression proposed by MacNeilage and Davis (1990a) …” (p. 214). Her conclusion from an analysis of these errors is that “The discussion … regarding the developmental pathways taken by children towards the learning of phonology is completely consistent with MacNeilage and Davis’s (1990a) model of language acquisition.” (p. 219).

At this point it is necessary to address the question of whether phonemes are indeed discrete, as Abler claims, in a way that the elements of chemistry and the chemical bases of inheritance are. The claim that phonemes are discrete is substantiated by speech errors that show that single segments are independently controllable in otherwise correct sequences. In addition to exchange errors there are 4 other types of errors in which segments act as independent entities (Shattuck-Hufnagel, 1979; MacNeilage, 2008, pp. 81-83); Shift errors: A unit disappears from its correct location and appears in another location—e.g. State loaned and (owned land): Addition errors: an extra unit is added to an intended sequence—e.g. spublic speaking (public): Omission errors: a unit is dropped from an intended sequence—e.g. sonata umber ten (number): Substitution errors: One intended unit is replaced by another—e.g. it’s a shallower test (chest).

Evidence of a Relation Between Lipsmacks and Babbling.

A key evolutionary hypothesis of the F/C theory is that syllables evolved from lipsmacks. Important behavioural evidence in support of this claim has recently been provided by Morrill et al., (in press) in a paper entitled “Monkey Lipsmacking Develops Like Speech”. In a developmental study of the dynamics of lipsmacks in macaque monkeys they explicitly tested the F/C hypothesis that syllables evolved from lipsmacks. They report that neonatal monkeys produce lipsmacks at the rate of about 3 per second, a rate similar to that of syllable production in infant babbling (Dolata, Davis and MacNeilage, 2008), while adult monkeys produced lipsmacks at the rate of about 5 per second, a rate similar to that of syllable production in adult speech (Dolata, Davis and MacNeilage, 2008). They also found that the chewing rate in these animals was initially slower than the neonatal lipsmack rate (about 2 per second), and this chewing rate did not change very much with development. The similarity of these repetition rates for neonatal and adult lipsmacks on the one hand and babbling and speech on the other is highly suggestive evidence that speech could have evolved from lipsmacks.

Comparative Neurobiological Evidence for the Syllable Frame.

It is well known in the comparative neurobiology of action that the production of biphasic cycles in general is controlled by neurophysiological entities known as a “Central Pattern Generators” (CPGs) A CPG is defined as “any network within the CNS that coordinates a motor behaviour or part thereof” (2006, p. 751). Grillner notes that “We know in all animals, vertebrates or invertebrates, movements are controlled by CPG networks that determine appropriate sequences of muscle activation. Each animal is endowed with a broad repertoire of CPGs located in different regions of the central nervous system, and available for differential activation, thus providing animals with a distinctive set of solutions to accommodate their widely divergent patterns of behaviour” (2006, p. 751).

There are two main categories of non-communicative voluntary action in vertebrates, one related to locomotion, and the other to food attainment and ingestion. With regard to locomotion, CPGs generate biphasic oscillatory cycles (e.g., for walking, swimming, and flying). Georgopoulos and Grillner (1989) point out that with the advent of terrestrial mammals, an increasing ability to control exact foot preference is superimposed on basic locomotory cycles. They also propose that in primates “the precise forelimb movements used to position the limb at will and to grasp different objects are very similar to, and have evolved from, those used to position the limb accurately during locomotion” (p. 1210).

With respect to ingestion, one can envisage a similar progression from a basic cycle to a modulation of that cycle. As part of selection for increasing food processing and ingestion in newly evolving warm-blooded mammals about 200 million years ago, an oscillatory CPG controlling the mandible for chewing, sucking, and licking may have developed, and in omnivorous forms ancestral to humans, an increasing ability to modulate the biphasic cycle for special processing purposes, especially those related to chewing, has followed. Most specifically, Lund and Kolta (2006) have proposed that brainstem circuits associated with the CPG that controls mastication “also participate in the control of human speech” (p. 381).

Historically, the main brain region associated with the motor control of speech is an area of the posterior inferior frontal lobe encompassing Broca’s area and its immediate surround. This area includes sites associated with the control of mastication in monkeys. Furthermore mirror neurons associated with lipsmacking have been found in area F5 of monkeys, which is considered to be the homolog of Broca’s Area in monkeys. Until recently, F5 had not been thought to be significantly involved in monkey vocalization. However, Coude et al. (2011) have recently recorded neurons in monkey F5 which discharged prior to and/or concurrent with the production of conditioned coo vocalizations in two monkeys.

Evidence for a specific role of syllable control in the posterior inferior frontal cortex of humans has been presented by Giraud et al. (2007). They have identified endogenous cortical rhythms which they associate with speech-related functions. Among other things they found a 3-6 Hz power band in the lower part of the motor cortex which, in their opinion, “offers a direct neural underpinning for the F/C theory of speech that assumes that syllables are phylogenetically and ontogenetically determined by natural mandibular cycles occurring at about 4 Hz.” (p. 1132). They consider that overall, their findings “emphasize the role of common cortical oscillatory frequency bands for speech production and perception and thus provide a brain-based account for the phylogenetic emergence and shaping of speech from available neural substrates" (p. 1133).

In addition to this possible involvement of posterior inferior frontal cortex, I have argued that with the evolution of speech from prespeech communicative capabilities, aspects of superordinate control of the mandibular cycle, or motor frame for speech moved from this region of frontal cortex to posterior medial frontal cortex. This control became mediated by a region formerly called the Supplementary Motor Area (SMA), but now divided into Pre-SMA and SMA subcomponents (MacNeilage 1998; MacNeilage and Davis 2001). The most highly developed neurobiological model of the control of speech production, a model which includes tests of quantitative predictions, has been presented by Bohland et al. (2010). They credit F/C theory for their adoption of posterior medial cortex as a site for frame generation.

This medial region was presumably responsible for the rhythmic repetitions of the same CV in Broca's famous patient "Tan", and similar repetitive CV automatisms in many other global aphasics, patients who lack the use of the lateral surface of the left hemisphere, while retaining the use of posterior medial frontal cortex (MacNeilage and Davis 2001). Electrical stimulation of this region, and the presence of irritative lesions affecting this region, also result in the production of automatisms in which a single CV form is rhythmically reiterated (MacNeilage and Davis 2001). These phenomena reveal that the basic rhythmic CV alternation capability, originating ontogenetically in babbling, remains present in the brain throughout the lifespan.

Finally, in recent years, comparative neurobiological evidence suggesting an evolutionary relation between lipsmacks and syllable frames comes from studies of the anterior insula. The insula is a region of cerebral cortex lying deep within the lateral sulcus, (sylvian fissure) intermediate between the superior medial surface and the inferior medial surfaces of the sulcus. The initial indication that the anterior insula might be involved in the evolution of speech involved a study of "Apraxia of Speech". Apraxia of speech (AOS) is "a syndrome assumed to reflect impaired higher order aspects of speech motor control, i.e. the "planning" of the respective vocal tract movement sequences" (Ackerman and Riecker, 2010, p 420). In a surprising finding in 1996, Dronkers reported that the only area of brain damage common to a group of N patients with apraxia of speech was the anterior region of the insula in the dominant hemisphere. Ackerman and Rieker (2010) summarize subsequent imaging studies that confirm a role of the anterior insula in the planning of speech production. They conclude that the region is "sensitive to phonetic-linguistic structure of verbal utterances" (p. 419), particularly emphasizing a study in which haemodynamic activity in the region is sensitive to the complexity of syllable structure.

Ackerman and Rieker observed that "considering … the well-established participation of the anterior insula in the phylogenetically rather old perceptual-motor functions such as feeding-related behavior … it is a bit bewildering that 'motor planning' of speech (Dronkers, 1996) or 'motor articulatory planning' (Nestor, 2003), capacities unique to our species, might might depend on the rostral, i.e. 'paralimbic' component of intrasylvian cortex." (p. 420) However, rather than being bewildering, it is what would be expected if one believes that the articulatory capabilities of speech evolved from food ingestion capabilities. Numerous imaging studies have shown involvement of the insula in both swallowing and chewing. The specific involvement of the insula in the motor action of human chewing, as distinct from effects of gustatory and olfactory input related to the chewed object, is indicated by a study in which electrical stimulation of the insula of epileptic patients produced chewing movements (Ostrovsky et al, 2000).

Interestingly, in the present context, these authors also noted the elicitation of "lipsmacks" in one patient. These might have been instances of pure frames unaccompanied by phonation. Thus, as might have been expected from the F/C perspective, human frames or framelike actions may have had a phylogenetic precursor in insular cortex. In a further confirmation of a phylogenetic relation between lipsmacks and syllables, Caruana et al. (2011) have elicited lipsmacks in monkeys by stimulation of the insula. These lipsmacks were dependent on social context, occurring only when the monkey was face to face with another individual, a circumstance that could presumably have been carried over into humans. The fundamental role of lipsmacks in monkey communication is emphasized by the finding of Ferrari et al (2009) that neonatal monkeys exchange lipsmacks with their mothers at birth. It may also be of interest, in the light of the possibility that lipsmacks derived from chewing, that the lipsmacks evoked by Caruana et al. sometimes terminated in chewing movements.

Subphonemic Particles?

Are phonemes the minimal particles of speech, or are they composed of smaller units? In the view of Studdert-Kennedy (1998), "consonants and vowels are not primitive units, but compounds, analogs of the molecule not the atom" (p. 209). There have been two candidates for the supposed subphonemic units of speech, the "distinctive feature" and the "gesture". A "distinctive feature" is an attribute of a segment that distinguishes the segment from another. For example the word "ten" is distinguished from the word "den" by a distinctive feature of voicing. The distinctive features are considered by the modern discipline of generative phonology to constitute a finite set of innate abstract entities out of which sound systems of languages are made. (See MacNeilage, 2008, pp. 225-235 for discussion.) Phonemes are considered to be bundles of distinctive features. But as Studdert-Kennedy (1998) points out, features "have no substance" (p. 210). As is evident from the adjectival terminology of all feature theories (nasal, coronal, continuant etc.) features are not independent entities (like atoms) but properties or attributes of the segments they describe. They are not action components. Instead, as Ladefoged (2006) points out, they "are best regarded as artifacts that linguists have devised in order to describe linguistic systems" (p. 12).

An alternative subsegmental action unit called the "gesture" has been proposed by members of a subdiscipline derived from dynamical systems theory called Articulatory Phonology, (e.g. Goldstein and Fowler, 2003) and endorsed by Studdert-Kennedy (1998; 2000). These gestures are actions involving 6 putative basic particles or "atoms" of speech, also called "organs". These particles/atoms, supposedly associated with 6 speech "organs" are the lips, the tongue tip, the tongue body, the tongue root, the soft palate and the glottis (Goldstein, Byrd & Saltzman, 2006). The gestures are considered to be oscillators, each involving the making and releasing of a vocal tract constriction. The "atoms" are considered to be combined into "molecules" (consonantal and vocalic forms) by naturally occurring contingencies arising from the coupling of oscillators, and influenced in their serial organization by a dimension of articulatory compatibility/incompatibility between pairs of gestures. (Goldstein, Byrd & Saltzman, p. 217)

The main problem with this conception is that, as we will now see, the use of the metaphors of "atom" and "molecule" is quite spurious. Atoms are the basic particles of matter and they combine into molecules to form matter. There are 118 atoms (called elements) of which 98 occur naturally, and they can be arranged consecutively in the Periodic Table according to the atomic weights indicated by their numbers (1 to 118). The individual structures are known in detail and precise rules for the way they can combine into molecules are also known (see Wikipedia—"Periodic Table" for discussion).

Nothing remotely resembling the state of affairs regarding atoms and molecules is true of the subsegmental/segmental level of speech production. The closest one can come to an analogy between chemical units and articulators is in the case of the soft palate, as it can function independently of other articulators. It typically makes a single binary contribution to speech (nasal, non-nasal) in a manner similar to the way that an element might participate or not participate in a molecule. But it clearly is different from the elements of chemistry because it retains its quality when combined, in a way that elements (e.g. oxygen in water) do not. Beyond that, the lower component of the lips, and the tongue are functionally dependent on the mandible, and are therefore not independently definable from an articulatory standpoint. Most importantly, there is no justification for regarding the three regions of the tongue (tip, body and root) as separate organs, across the board. The common sense conclusion that the tongue is not three organs but one is the correct conclusion. The main manifestation of its unity as an organ is the fact that one of its muscles, the genioglossus muscle, originating at the chin, fans radially into the entire tongue. Although it is not necessarily always activated as a whole, there is no evidence that it has 3 separate functional subdivisions. Why are there not 4 organs? Why is the tongue blade, often a separate component in descriptions of the tongue, not also included as one of the "atoms" of speech? And why is the mandible not considered to be an "atom" of speech, given that it is clearly a discrete organ anatomically, and it is almost as important in the formation of constrictions, and their releases, as the tongue is? In short, with the exception of the soft palate, where discreteness is self evident, discreteness of the putative subsegmental gestures has not been established.

For a gestural perspective to be successful, it should be able to characterize the different kinds of consonants and vowels in languages. But in contrast to the realm of generative phonology, in which the unit "distinctive feature" is roughly co-extensive with the gesture, (see MacNeilage, 2008 for discussion) such a comprehensive characterization has never been attempted. A fundamental issue here is that finiteness entails discreteness. One can't count entities unless they are discrete, and, with the exception of the soft palate, the discreteness of gestures is not evident. But beyond this, consider the problem at an empirical level: in a count of the segmental content of 317 languages Maddieson (1994) found a total of 558 consonants and 210 vowels. The main contributor to this huge number of segments is tongue position. How might these individual sounds be constructed from the 6 "atoms" of articulatory phonology, including only 3 "atoms" involving the tongue? Not only is no attempt made to address this problem, but conception of the tongue as three organs is not even used in the approaches of the AP group to either phylogeny or ontogeny, beyond the assertion that these organs exist.

Explaining the Particulate Principle

How should the common manifestation of the particulate principle in the domains of chemical interaction, inheritance and language be explained. According to Abler "the several particulate systems derive their properties by common inheritance from …. the particulate principle of self-diversifying systems which predates them all." (p. 1) Abler believes that the principle itself derives from "some underlying property of material nature." (p. 12). By material nature he means some property of the physical world. This property is considered to exist a priori, as in the "essences" of Plato. (See MacNeilage, 2008 for a discussion of essentialism with reference to language in particular.). "Particulation" is given axiomatic status as "a Euclidean principle which acts in a universe governed by classical arithmetic" (p. 12). As to the origin of the phoneme in particular, it is considered to represent "a class of discrete mental categories which probably depends, at some level, on the discrete, or particulate, nerve impulse…." (p 12).

It can be seen that Abler has a reductionistic view of physics as the basis of lawfulness in other domains.

In spite of the conviction that all sciences ought to be based on the same principles because all are based on physics, scientists use different systems of rules in dealing with various natural systems, implicitly treating such systems as if they were unrelated. The discrepancies are nowhere greater than between social, biological and physical sciences.

It is obvious that the present perspective is somewhat different from Abler's perspective. In general terms, it derives from Crick's contention that the mechanism of natural selection "makes biology different from all the other sciences." (Crick, 1988, p. 5) As the phoneme evolved by natural selection, it cannot be explained by the laws of physics. Instead, it evolved via a typical mode of selection—it evolved by tinkering (Jacob, 1977). First, the central pattern generator for ingestion was tinkered into use for vocal communication in the form of the syllable frame. Then the two phases of the frame cycle were tinkered into use as programmable slots enabling the differentiation of linguistic messages by means of phonemes. This mode of origin is different in its overall form from the origin of the particulate principle in both the domain of chemical interaction and in the domain of inheritance. The phoneme emerged from a superordinate structure. Consideration of the origins of the chemical elements (Penzias, 1978) and the 4 chemical bases of inheritance, Freeland and Illardo (in press) shows that these particles were synthesized from subcomponents— elements from protons and neutrons and electrons, and the nucleic acids constituting DNA and RNA, the two substances associated with the genetic code, from elements.

Thus, in the present view, the basis for the evolution and development of phonemic systems is a pre-systematic frame. In the case of the labial consonants and central vowels of pure frames, the emergence of the systematic basis of phonemes involves the provision of active control of articulatory positions which were previously produced passively, as inevitable side-effects of the two phases of the frame cycle. In the case of lingual frames the main task is to constrain tongue movements in the time domain to more or less coincide with the durations of the closing phase for consonants and the opening phase for vowels.

References

Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures, 12*: 1-13.

Ackermann, H. & Riecker, A. (2010). The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Structure and Function, 214*: 419-433.

Bohland, J.W., Bullock, S. & Guenther, F.H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*: 1504-1529.

Breen, G. & Pensalfini, R. (1999) Arrernte: A language with no syllable onsets. *Linguistic Inquiry, 30*; 1-26.

Caruana F., Jazzini A., Sbriscia-Fioretti B., Rizzolatti G. & Gallese V (2011) Emotional and socialbehaviors elicited by electrical stimulation of the insula in the macaque monkey. *Current Biology 21*: 1-5.

Crick, F. (1988). *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books.

Davis, B. L.& MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*: 1199-1211.

Davis, B. L. & MacNeilage, P. F. (2002). The internal structure of the syllable. In T. Givón & B. F. Malle (eds.), *The Evolution of Language out of Prelanguage*. Amsterdam: John Benjamins, 135-154.

Dolata, J., Davis, B.L. & MacNeilage, P.F. (2008). Characteristics of the rhythmic organization of babbling: Implications for an amodal linguistic rhythm. *Infant Behavior and Development 31*: pp. 422-431.

Donald, M. (1999). Preconditions for the evolution of protolanguages. In M.C. Corballis & S.E.G Lea (eds.), *The Descent of Mind: Psychological Perspectives on Hominid Evolution.* Oxford: Oxford University Press, 138-154.

Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. Nature: 384, 159-161.

Dunbar, R.I.M. (1996). *Grooming, Gossip and the Evolution of Language.* Cambridge, MA: Harvard University Press.

Ferrari, P. F., Gallese, P., Rizzolatti, G. & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth movements in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*: 1703-1714.

Ferrari, P.P., Paukner, A., Ionica, C. & Suomi, S.J. (2009). Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Current Biology 19*: 1768-1772.

Freeland, S.J. & IIlardo, M. (In press). Language evolution in humans and ancient microbes: What can human language acquisition tell us about the origin of genetic information?

Gazzaniga, M. & Heatherton, T.F. (2003). *Psychological Science (2nd Edition)* New York: Norton.

Georgopoulos, A. P. & Grillner, S. (1989). Visual-motor coordination in reaching and locomotion. *Science, 245*: 1209-1210.

Giraud, A-L., Kleinschmidt, A., Poeppel, D., Lund, T. & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron, 56*: 1127-1134.

Goldstein, L.M. & Fowler, C.A. (2003). Articulatory phonology: A phonology for public language use. In N. Schiller & A. Meyer (eds). *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* Berlin: Mouton de Gruyter, 159-207.

Goldstein, Louis M., Byrd, D. & Saltzman,E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M. (ed.), *Action to Language Via the Mirror System.* Cambridge: Cambridge University Press, 215-249.

Grillner, S. (2006). Biological pattern generation: The cellular and computational logic of networks in motion. *Neuron, 52*: 751-766.

Humboldt, W. von (1836). Uber die Verscheidenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwichelung des Menschengeschlechts. Berlin: Royal Academy of Sciences. Republished (G.C. Buck and F. Raven, trans.) (1971) *Linguistic Variability and Intellectual Development.* Baltimore, Maryland: University of Miami Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.

Jacob, F. (1977). Evolution and tinkering. *Science, l96*: ll6l-1166.

Jaeger, J.J. (2005). *Kid's Slips: What Young Children's Slips of the Tongue Reveal About Language Development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium.* New York, NY: Wiley, 112-136.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 48*, 1-22.

Lindblom, B. & MacNeilage, P.F. (2011). Coarticulation: A universal phonetic phenomenon with its roots in deep time. FONETIK 2011, 41-44 In – *QPSR TMH Vol 51*, KTH Stockholm, http://speech.kth.se/prod/publications/files/3588.pdf

Lund, J.P. & Kolta, A. (2006). Brainstem circuits that control mastication: Do they have anything to say during speech? *Journal of Communication Disorders, 39*: 381-390.

MacNeilage, P. F. (1985). Serial ordering errors in speech and typing. In V. A. Fromkin (Ed.), *Phonetic Linguistics.* New York, NY: Academic Press: 193-201.

MacNeilage, P. F. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*: 499-548.

MacNeilage, P.F. (2008). *The Origin of Speech.* Oxford: Oxford University Press.

MacNeilage, P.F. (In press). Sound Patterns of First Words and How They Became Linked With Concepts. In Lefebvre C (ed.) *The Origin of Language.* Cambridge University Press, Cambridge.

MacNeilage, P. F. & Davis, B. L. (1990a). Acquisition of speech production: Frames, then content. In M. Jeannerod (ed.), *Attention and Performance X111: Motor Representation and Control.* Hillsdale, N.J.: Erlbaum, 453-476.

MacNeilage, P.F. & Davis, B.L. (1990b). Acquisition of speech production: The achievement of segmental independence. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer, 55-68.

MacNeilage, P. F. & Davis, B. L. (2000). On the origin of the internal structure of word forms. *Science, 288*: 527-531.

MacNeilage, P. F., & Davis, B. L. (2001). Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology, 11*: 696-700.

MacNeilage, P. F. & Davis, B.L. (2002). The origins of intersyllabic complexity. In T. Givón & B. F. Malle (eds.), *The Evolution of Language out of Prelanguage.* Amsterdam: John Benjamins, 135-154.

MacNeilage, P.F. & De Clerk, J.L. On the motor control of coarticulation in CV syllables. *Journal of the Acoustical Society of America, l970: 45*, 1217-1233.

MacNeilage, P. F. Davis, B. L., Kinney, A. & Matyear, C. L. (1999). Origin of serial output complexity in speech. *Psychological Science, 10*: 459-460.

MacNeilage, P. F., Davis, B. L., Kinney, A. & Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. *Child Development, 71*: 153-163.

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Morrill, R.J., Paukner, A., Ferrari, P.F. & Ghazanfar, A.A. (In press). Monkey lip-smacking develops like human speech. *Developmental Science.*

Nestor, P.J., Graham, N.L., Fryer, T.D., Williams, G.B. Patterson, K. & Hodges, J.R. (2003). Progressive non-fluent aphasia is associated with hypometabolism centered on the left anterior insula. *Brain, 126*: 2406-2418.

Ostrowsky, K., Isnard, J., Ryvlin, P., Guenot, M., Fischer, C. & Mauguiere, F. (2000). Functional mapping of the insular cortex: clinical implication in temporal lobe epilepsy. *Epilepsia, 41*: 681-686.

Penzias, A.A. (1978). The origin of elements. (Nobel Lecture) http://nobelprize.org/nobel_prizes/physics/laureates/1978/penzias-lecture.pdf

Redican, W. K. (1975). Facial expressions in nonhuman primates. In L. A. Rosenblum (ed.), *Primate Behavior: Developments in Field and Laboratory Research, Vol. 4.* New York: Academic Press, 103-194.

Rousset, I. (2003). From lexical to syllabic organization: Favored and disfavored co-occurrences. *Proceedings of the 15th International Congress of Phonetics.* Barcelona: Autonomous University of Barcelona, 2705-2708.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum, 295-342.

Studdert-Kennedy, M.G. (1998). The particulate origins of language generativity: From syllable to gesture. In Hurford, J. R., Chris Knight, C. & Studdert-Kennedy, M.G. (eds.), *Approaches to the evolution of language.* Cambridge: Cambridge University Press, 202-221.

Studdert-Kennedy, M. G. (2000). Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In Knight, C, Studdert-Kennedy, M.G. & Hurford, J.R. (eds.), *The evolutionary emergence of language: Social function and the emergence of linguistic form,* 161-176. Cambridge: Cambridge University Press.

Van Hooff, J.A.R.A.M. (1967). Facial displays of the catarrhine monkeys and apes, in Morris, D. (ed.), *Primate Ethology.* London: Weidenfield and Nicholson, 7-68).

On the open secret of language[*]

Jan Anward

Linköping University

jan.anward@liu.se

Abstract

Language is not inside speaking but comes forth as a consequence of dialogue and tradition.

Speaking is independent of language

In our halcyon days, some of us were intrigued by the idea of a language machine, sitting in our heads, and cranking out well-formed sentences for us. Linguistics was all about reconstructing the design of that machine. And early psycholinguistic studies (see Greene 1972 for an overview) seemed to corroborate that view, managing to link design features of the machine directly to mental processes.

However, the language machine did not last long. McCawley (1968) fairly soon showed that well-formedness did not result directly from any productive device, but must be understood as conditions on the output of such a device. Grammar, if anything, is a filter on independent productions. And the short-lived experiment of Generative Semantics conclusively demonstrated that well-formedness, if at all a meaningful notion, is inextricably context-dependent (Lakoff 1989).

The obvious conclusion is that humans have a capacity for a particular mode of sound production, speaking, which is appropriated in doing language, but which is not necessarily constrained by established linguistic norms. Activities such as babbling (Lindblom 1997), glossolalia (Samarin 1972, Holm 1996), and sound poetry (Keith 2005, Olsson 2005: ch. 6), where speaking does not adapt to recognized norms, provide further support for this conclusion. It also fits very nicely with MacNeilage's carefully elaborated argument (MacNeilage 2008) that speaking is itself an appropriation of existing action schemata, and evolutionarily independent of language.

Speaking is the iteration of syllables, grouped into intonation units (see Szczepek Reed 2010 and Anward 2010 for a recent discussion), as in (1) below, an extract from a telephone conversation.

(1) >hejdeee:va.

hardumammahemma¿<

(>hi-it-is-Eva.

have-you-mother-home¿<)

---

[*] This is for Björn Lindblom, inspiring partner in many conversations, and prime mover in the quest for a derivation of language from non-language. This article is a kind a précis of a much longer work in progress, called *Doing language.*

In the first line of (1), there are five syllables. One of these, the fourth one, is carrying pitch accent (marked by underlining) and they are bound together by a falling intonation contour (marked by the final period). In the second line, there are six syllables, a pitch accent, and a slightly rising intonation contour (marked by the final inverted question mark).

Language comes from dialogue

So where does language come from? A language, a *langue* in a modified Saussurean sense (Saussure 1916, particularly ch. 4), is a regular outcome of conversation. Conversation is grounded in a natural attitude of humans, a dialogical attitude (Linell 2009), which compels humans to hear a sound as an address and an aboutness, to answer an address with a returned address and a continued aboutness, and to assume that other humans work in the same way as you do.

Participants informed by a dialogical attitude will accomplish activities where they take turns at connected dialogical actions, and where each action is relevant to the one preceding it, and anticipates a relevant action following it. And these dialogical actions serve to update an ongoing social activity, by introducing into it and maintaining there participants and relations among them. Some of these participants and relations are physically present in the activity ('real'), while others are only symbolically present, i.e. they enter and remain in the activity only through linguistic expression.

The result is a conversation, or a dialogue, as in (2), which is the opening of a land-line phone call (recorded by Anna Lindström; see Lindström 1994). Eva calls a neighbour in a practical matter. Bodil, a young girl who is visiting the neighbours' house to play with their daughter Veronika, answers the phone.

(2) Pippiperuk 1, opening

00 ((fyra signaler))
01 Bodil: >sexsjutvåfyra¿<
02 Eva: >hejdeee:va.
03 hardumammahemma¿<
04 Bodil: näe:.

(00 ((four signals))
01 Bodil: >six-seven-two-four¿<
02 Eva: >hi-it-is-Eva.
03 have-you-mother-home¿<)
04 Bodil: noo:.)

The first two turns of (2) is a classical example of an adjacency pair (Schegloff & Sacks 1973, Schegloff 2007), where, in the analysis of Schegloff (1968), a first action, a summons (line 00), makes a second action, an answer (line 01), relevant, and where the pair as a whole makes further talk (line 02) relevant. The answer to the summons is not just any answer, though, but an identification, which is accomplished through recital of a telephone number, and this identification, in its turn, makes a second identification relevant. The second identification, by first name this time, follows in the third turn (line 02), and it is preceded by a greeting, which makes a second greeting relevant. However, Eva does not wait for a second greeting. Instead, she asks Bodil to pass on the phone call to an adult, thereby deferring the second greeting by means of an inserted question.

In accomplishing this small stretch of interaction, Eva and Bodil also do a small *langue*. Their unfolding responsory of dialogical actions simultaneously unfolds a system of syntagmatically related turns, where each turn is demarcated by a change in voice.

The ensuing system looks like this:

((fyra signaler))		summons		
Bodil:	>(sexsjutvåfyra)¿<	answer	identification	
Eva:	>hejdeee:va.		identification	greeting
	hardumammahemma¿<			question
Bodil:	näe:.			answer

In this way, by taking turns at dialogical actions, participants continuously structure their ongoing conversation like a language (to borrow a formulation from Lacan 1977).

And from tradition

And when participants speak in a tradition, that is, model new conversations on old ones, they provide their conversations with further structuration.

After the phone call, whose opening we saw in (2), Eva makes two more calls, to two other neighbours. The openings of these calls follow in (3) and (4).

(3) *Pippiperuk 2, opening*

00 ((två signaler))

01 Karin: två noll noll sju två¿

02 Eva: ja hejsan de va >eva andersson<,

03 (0.3)

04 Karin: m[hm¿

(00 ((two signals))

01 Karin: two zero zero seven two¿

02 Eva: yes hi-*san* it was >Eva Andersson<,

03 (0.3)

04 Karin: m[hm¿)

(4) Flintskalleperuk, opening

00 ((två signaler))

01 Sylvia: sylvia¿hh

02 Eva: ja hejsan de va >eva andersson:<

03 Sylvia: ja hej:

(00 ((two signals))

01 Sylvia: Sylvia¿hh

02 Eva: yes hi-*san* it was >Eva Andersson:<

03 Sylvia: yes hi:)

It is plain to see that (3) and (4) are versions of (2).

In all three cases, a summons (line 00) is followed by an answer, which is also a first identification (line 01). Then comes a first greeting, followed by a second identification (line 02). And then in (3), like in (2), the second greeting which is made relevant by the first greeting does not directly come forth, but in (4), it does.

But there are also differences between the two openings. The third turn of *Pippiperuk 1* consists of two intonation units:

`hejdeee:va.`

and

`hardumammahemma¿<`

Through her first intonation unit, Eva accomplishes a first greeting and a second identification. Through her second intonation unit, she cancels the relevance of a second greeting by her current interlocutor.

In constructing the third turn of *Pippiperuk 2*, Eva uses only one intonation unit:

`jahejsandeva>evaandersson<,`

Again, she does a first greeting and a second identification, but this time she does not cancel the relevance of a second greeting. This time, Eva is not talking to a child, who she does not want to talk to, she is talking to another adult, a neighbour who not necessarily will recognize her (and in fact does not at first) and who she does want to talk to.

This difference in social relations between the two contexts is indexed by formal differences between the two turns: a more elaborated first greeting, where *ja* (yes) and the suffix *san*[1] are added to *hej* (hi), and a more formal second introduction, which includes both Eva's first name and her last name, and where the tense is changed, from present (*e*; is) to past (*va*; was).

The opening of *Flintskalleperuk* is cast in the same mold, and differs from that of *Pippiperuk 2* only in the placement of the pitch accent in the third turn - and, in the occurrence of a second greeting.

Thus, the openings of these three phone calls resonate[2] with each other, in that their articulations into turns and intonation units match quite closely. When we align recurrent, i.e. formally and functionally similar, units vertically, in a diagraph format (Du Bois 2004), we see how the openings through this resonance are integrated into one system, where a further dimension of difference, Jakobsons's axis of selection (Jakobson 1956: 60), comes into play. This is shown in (5) below.

Dialogical actions are indicated in abbreviated form (a = answer, g = greeting, i = identification, q = question, r = reaction, s = summons).

---

[1] The suffix san is derived from the parenthetical sa'n 'said-he', and now serves to strengthen certain interjections.

[2] For this notion of resonance, see Anward & Lindblom 1999: 14, 27, and 37, and Du Bois 2004.

(5) Pippiperuk 1, Pippiperuk 2, and Flintskalleperuk, openings integrated

```
 ((fyra signaler))                                            s
           B: >(sexsjutvåfyra)¿<                              a i
                       E: >hejdeee:va.                        i g
                                hardumammahemma¿<             q
                                             B: näe:.         a


((två signaler))                                              s
           K:    tvånollnollsjutvå¿                           a i
                       E: jahejsandevaevaandersson,           i g
                                             (0.3)  K: mhm¿   r


((två signaler))                                              s
         S: sylvia¿hh                                         a i
                       E: jahejsandevaevaandersson:           i g
                                             S: jahej         g
```

Any such diagraph can be interpreted as a *langue*, that is, a network, or system, of syntagmatically, paradigmatically, and associatively[3] related units. Relations of temporal precedence translate straightforwardly into syntagmatic relations, and vertically aligned items are paradigmatically related.

The resonance of the openings also affords further articulation of them, through the application of the following principle, the fundamental 'discovery procedure' of structural linguistics (see e.g. Gleason 1961, chs. 5-7; and also Peters 1983: 37, and Anward & Lindblom 1999: 28).

(6) Segmentation

A turn which contains a recurrent part is articulated into that part, a preceding environment, and a following environment.

The recurrent items *två* and *sju* articulate the second turns of (2) and (3) into *sex två sju fyra* and *två nollnoll sju två*, and the recurrent items *hej*, *de* and *eva* articulate the third turns of (2), (3), and (4) into *hej de e eva hardumammahemma* and *ja hej san de va eva andersson*.

---

[3] Associative relations, relations of similarity, in form or meaning, are the system-defining relations introduced by Saussure (1916 [1967]: 170-175) alongside syntagmatic relations. Hjelmslev (1959: 152) proposed to substitute paradigmatic relations for associative relations, but in fact, all three types of relation capture essential features of conversational practice.

And then we arrive at the langue in (7).

(7) Pippiperuk 1, Pippiperuk 2, and Flintskalleperuk, openings integrated

```
 ((fyra signaler))                                               s

          B: >(se̲x sju två̲ fyra)¿<                               a i

                    E: >hej de e e̲:va.                           i g
                            hardumma̲mmahemma¿<                   q

                                        B: näe:.                 a


((två signaler))                                                 s

          K:   två̲ nollnoll sju två̲¿                             a i

                    E: ja he̲j san de va eva andersson,           i g

                                        (0.3) K: mhm¿            r


((två signaler))                                                 s

        S: sylvia¿hh                                             a i

                    E: ja hej san de va eva a̲ndersson:           i g

                                        S: ja hej                g
```

By taking turns at dialogical actions, and by modelling new conversations on old conversations, participants continuously structure their conversations like a language, along the axes of combination and selection, and down to the level of words and phrases.

References

Anward, J. (2010). Making units: Comments on Beatrice Szczepek Reed "Intonation phrases in natural conversation: A participants' category?". Barth-Weingarten, D. et al. (eds.). *Prosody in interaction.* Benjamins. 213-216.

Anward, J. & Lindblom, B. (1999). On the rapid perceptual processing of speech: From signal information to phonetic and grammatical knowledge. http://www.liu.se/ikk/medarbetare/jan-anward/filarkiv-publikationer/1.209284/ja-1999b-anward-lindblom.pdf

Du Bois, J. W. (2004). Towards a Dialogic Syntax. Ms.

Gleason, H. A. (1961). *An Introduction to Descriptive Linguistics*. Holt, Rinehart & Winston.

Greene, J. (1972). Psycholinguistics. Chomsky and Psychology. Penguin.

Hjelmslev, Louis, 1959: *Essais linguistiques*. Travaux de Cercle Linguistique de Copenhague XII.

Holm, N. G. (1996). Extas, trans och tungotal. Cullberg, J. et al. (eds.). *Mänskliga gränsområden.* Natur och kultur. 35-59.

Jakobson, R. (1956). Two aspects of language and two types of aphasic disturbances. Jakobson, R. & Halle, M. *Fundamentals of language*. Mouton & Co.

Keith, Th. (2005). Poetische Experimente der deutschen und russischen Avantgarde (1912-1922) – ein Vergleich. Weidler.

Lakoff, R. (1989). The way we were; or; the real actual truth about generative semantics: a memoir. *Journal of Pragmatics* 13. 939-988.

Lindblom, B. (1997). Talet tar form. De Geer, B. et al. (eds.). *Från joller till läsning*. Gleerups. 11-32.

Lindström, A. (1994). *Language as social action.* Skrifter utgivna av Institutionen för nordiska vid Uppsala universitet 46.

Linell, P. (2009). Rethinking language, mind, and world dialogically. Information Age Publishers.

MacNeilage, P. F. (2008). *The Origin of Speech.* Oxford University Press.

McCawley, J.D. (1968). Concerning the Base Component of a Transformational Grammar. *Foundations of Language* 4. 243-269.

Olsson, J. (2005). Alfabetets användning: konkret poesi och poetisk artefaktion i svenskt 1960-tal. OEI editör.

Peters, A. M. (1983). *The Units of Language Acquisition.* Cambridge University Press.

Samarin, W. J. (1972). Tongues of men and angels: the religious language of pentecostalism. Macmillan.

Saussure, F. de (1916). *Cours de linguistique générale*. Payot.

Schegloff, E. A. (1968). Sequencing in Conversational Openings. *American Anthropologist* 70. 1075-1095.

Schegloff, E. A. & Sacks, H. (1973). Opening Up Closings. *Semiotica*VIII. 289-327.

Schegloff, E. A. (2007). *Sequence organization in interaction.* Cambridge University Press.

Szczepek Reed, B. (2010). Intonation phrases in natural conversation: A participants' category? Barth-Weingarten, D. et al. (eds.). *Prosody in interaction.* Benjamins. 191-212.

Are Natural Languages Codes?

Per Linell

Department of Culture and Communication, Linköping University[4]

per.linell@liu.se

Abstract

Is a natural language a code, that is, a system for mapping one kind of representation onto another in a rule-governed manner? Or more precisely, does each language contain (a set of) codes? This paper contains a brief discussion of this issue with regard to three levels, pragmatics (utterances in relation to situated meanings), formal grammar (syntax in relation to formal semantics), and phonology (the nature of the "phonemic code", and the relations between sounds and writing, or vice versa). The answers given are, in a nutshell, "no", "more or less no" and "perhaps (although in a limited sense)". Finally, the paper briefly discusses if analogies can be drawn from language to the so-called genetic code.

Keywords: language; code; genetic code; situated meaning; formal syntax; compositionality; phonology; phonological gesture; speech, writing.

Introduction

Roman Jakobson (1972) proposed the "code", that is, the language system, as one out of six basic aspects of human communication. Like most other linguists, he saw language (= the code) as a set of conventions used to communicate meaning. As a first approximation, such a formulation is probably acceptable to most scholars of language. But in many contexts, a more precise definition of code will be necessary. In this paper I shall begin to probe the issue if, and in what sense, if any, a natural language (or some part of it) is a code.

A few words about my general starting-point: One may take the fundamental nature of language to be either *abstract forms* organised as systems (in phonology, syntax or semantics), or *actions* and processes (or patterns of such processes) that are part of situated sense-making practices. The theories developed within mainstream linguistics give priority to the first-mentioned perspective. In the old days, structuralist phonologists used to talk about some of this in terms of the priority of "form" over "substance" (embodied performance, e.g. in phonetic "realisations") (e.g. Fischer-Jørgensen 1975). I want to reverse the priorities. A dialogical perspective, i.e. an approach that emphasises interactions and context-interdependences, must give priority to actions, processes and practices (Linell, 2009). In adopting this stance, I prefer the term "languaging" to the more common one of "language use", since the latter term clearly implies that the language system has priority over the use (Linell, 2012a). In accordance with this, Anward (2011) recommends that we start with how people are "doing language", which would then comprise both the situated languaging and the emergent language that results from the practices of languaging. Similarly, Lindblom (this volume) adopts what he calls a "performance-based" perspective. He points out that this amounts to a non-structuralist approach, thus opposing the mainstream of modern

---

[4] Also Department of Education, Communication and Learning (IPKL), Göteborg University.

linguistics (Saussure, Chomsky etc. and all their followers who have assumed that language structure must necessarily have a logical priority over performance, or in our terms: languaging). Admittedly, the formal perspective cannot be abolished altogether within a theory of languaging (see below, and Linell, 2013), but dynamics and interactions are more basic.

Is (a) language a code? I will discuss this with regard to three levels, pragmatics (the relation of utterances to situated meanings), formal grammar (the relation of syntax to formal semantics), and phonology (the relation of sounds to graphic representations, and vice versa). After some introductory remarks on the notion of "code", I will discuss compositionality in language, which is one kind of tack on the position that language is a code. After that, I shall pay particular attention to phonology, and raise some issues pertaining to biology and culture, and to speech vs. writing and literacy. Finally, I shall add some brief remarks on the analogy between language and the so-called genetic code.

Codes

Key to most accounts of what one should mean by a "code" is its presumed function in the exchange of "information", especially cognitive information. The code would primarily be seen as consisting of discrete expression-meaning pairings, or as a set of (formally) specified conventions for mapping, or translating, one kind of linguistic representation or "(en)coding" (e.g. speech) onto another (e.g. writing, that is, speech encoded in phonemes is re-encoded into graphic signs), or from one symbolic representation (linguistic expression) to semantic or pragmatic interpretations ("meanings"). The relations involved in these different cases are not quite the same.

In a coding relation, the source representation should be relevant and sufficient for deriving the re-coded form. Such mappings could be unidirectional (from system A to system B) or bidirectional (between A and B), preferably in one-to-one, but sometimes in one-to-many, many-to-one or many-to-many relationships (one can think of such simple things as homonymy and polysemy of words). A canonical example of a one-to-one coding relation is that between alphabetic writing and Morse code letters (dots and spaces), or vice versa. Another example would be written words (strings of letters) translated into an ASCII code (strings of bits). Parts of mappings from syntax to formal meaning (or vice versa) or from speech to writing (or vice versa) might be a bit more of one-to-one (but see below). The relations of linguistic utterances to situated meanings, or vice versa, would normally involve many-to-many relations. In a formal(ised) model, one would ideally set up algorithms for mapping. However, it is rather obvious that the relations between linguistic expressions and their meanings are not pairings like alphabetic letter–Morse signal. Indeed, the relations are at best metaphorically code-like.

The point that languages are not strictly codes has been made several times (e.g. Kravchenko, 2007; Love, 2007). In this paper, I will only make some variations on this theme.

Compositionality

In linguistic theory, the thesis of compositionality has been a common assumption. Briefly put, this thesis states that the meaning of a complex expression can be entirely derived from the meanings of the constituent expressions and the ways in which these are combined (the abstract structure of the sentence of which constituents are part). Preferably, the derivation of compositional meaning should be computable by a formal calculus (Pagin 2003). Compositionality is a way of giving substance to the idea that a language is a code, that is, a set (more precisely: an infinite set) of expression-meaning pairings. I shall discuss the applicability of compositionality at two levels, pragmatics and formal semantics.

Pragmatics

Usually, compositionality is seen as a property of semantics, i.e. the semantic properties tied to linguistic expressions, most often sentences seen as linguistic types rather than as situated utterances. Pragmatics, however, deals with the situated meanings, sometimes called utterance meanings, that are communicatively relevant for parties to the particular sense-making events (I will use the abbreviation SCRM for ”situated communicatively relevant meaning”). The SCRM of a particular utterance or contribution to a communicative interaction usually involves imagining the speaker´s ”why of communication” (“why that now”; Bilmes, 1985); parties must understand how the particular utterance is intended, or could arguably be intended, as a contribution to a situated communicative project in progress. For example, an utterance like *I am married* will be interpreted in different ways in different situations, depending on how participants understand the reasons for the speaker saying precisely this. For example, if, in a situation of emergent attempts of seduction, the seducee utters *I am married*, there can be no compositionality principle mapping the semantic properties of the constituent words (and the syntactic structure) onto the possibly implied SCRM `There is no point in trying´. In real life, we go beyond the information given in the linguistic make-up (Bruner, 1973); “pragmatic enrichment” (Carston, 2002) takes place.

In other words, language itself is merely allusive and indexical, in relation to SCRMs. The why of communication is often not explicitly expressed in so many words; contextual factors and other semiotic resources, such as manual gestures and facial expressions, are important too. Moreover, constructions and their situated interpretations cannot be specified before the utterances have been (at least partially) completed, since they are “co-genetic” with contexts. That is, relevant contexts cannot be predicted beforehand; they are made relevant as part of participants´ communicative actions.

There are often differences between what is explicitly said and what is meant (SCRMs) in a situation. In some exceptional situation types (genres), however, there is a strive for a very close relation between linguistic expression and intended meaning, but this is the exceptional case, not the general principle. And even in a case such as law, the various legal rules must be interpreted in relation to specific cases; hence the need for a body of exemplary practical applications (“precedent cases”). One might also argue that a sentence in itself, such as *I am married*, taken in abstracto, has a decontextualised ”linguistic meaning”, but this meaning is hardly ever sufficient as an SCRM, except perhaps in games like exercises of logic or linguistic lessons, in which the purpose is to bracket (as completely as possible) considerations of normal SCRMs.[5]

Languaging in speech (including its non-vocal accompaniments) consists of embodied movements, gestures in time, with multimodal forms of expression. Activities of languaging are driven not only by cognitive incitements but also by interests, emotions, volitions, and of course environmental (“external) influences. They are dynamic movements, yet they are structured, and their progressions can therefore often be projected (Auer, 2005, 2009). However, projections can be changed in the course of (inter)action; they can be redirected, interrupted, restarted, and recycled with variations, due to e.g. impressions, sudden insights or whims, initiatives and counter-moves by interlocutors, affordances and obstacles of “external” objects and events (Linell, 2013). Linguistic resources (of language) must arguably be designed to fit these conditions of languaging. They must be accommodated to dynamics, the nature of embodied, temporal actions and movements etc.; plasticity, stretchability, fuzzy boundaries, and open potentialities to be exploited in contexts.

---

[5] The dialogical theory of communicative projects and activities, including a theory of situated meanings, is sketched in Linell (2009). It is important to note that, according to this theory, situated meanings (SCRMs) and ”linguistic (semantic) meanings” (tied to expressions in abstracto) are not phenomena of the same kind. The latter would preferably be seen as abstract *meaning potentials*, i.e. resources that can be used in combination with various contextual resources by participants in communicative (or cognitive) activities when they ”mean” something in situ (Norén & Linell 2007).

By way of summary, utterances cannot be assigned situated meanings solely on the basis of their linguistic composition. Thus, a principle of compositionality does not hold for situated meaning. In other words, in actual languaging, language (as a whole) *cannot be a code.* But what about the language system in a narrow sense? Does compositionality hold for the formal-semantic representations of sentences taken as linguistic entities? Let me briefly take up a few relevant aspects of this issue in the next section.

Formal syntax and structural semantics

This level is where the discussion of the compositionality of complex (syntactic) linguistic expressions usually takes place. Even if compositionality does not hold for the pragmatic level (SCRMs) (section "Pragmatics"), it could perhaps be valid for the abstract linguistic sentences, such as the sentence *I am married* as an abstract *type* in a formally defined *system of sentences*, rather than as a token in a situated act of languaging? Pagin (2003), for example, claims that compositionality, evidently at this level, is functional, partly because it would have an ev*olutionary value;* it is purportedly advantageous for human beings to have developed a communication system with compositionality, because this is more efficient than one without it.

Nobody would argue, I believe, that sense-making in complex activities is possible unless there is *some kind* of compositionality, at least in a loose sense. Yet, many, but hardly all, semanticists argue for at least some degree of context-sensitivity in linguistic semantics (semantics taken as dealing with the meanings tied to abstract expressions).[6] In my own opinion, it would be strange if these linguistic resources were not influenced by the openness of actual languaging. Strict compositionality would not necessarily be an evolutionary advantage, as Pagin (op. cit.) suggests; why would it be advantageous to be constrained by a rigid language in a world in which new types of communicative demands (i.e. demands that cannot be sufficiently pre-specified in abstracto) tend to turn up continuously in the life of the single individual (and communities too)? It should also be noted that successful communication does not imply that participants arrive at completely shared understandings of things talked about (e.g. Linell, 2009). Instead, we need only sufficient understandings "for current practical purposes". If serious problems of mutual understandings seem to come about, we may fortunately (although, alas, not in all social situations) appeal to strategies of repair, as has been amply demonstrated in Conversation Analysis.

The syntactic structures relevant to a theory of compositionality must be abstract and general. However, in actual communicative languaging the creation of SCRMs is primary, and here we are faced with incremental processes both in speech production and speech perception (Linell, 2012b, 2013); in these context-interdependent processes (contexts involving both local sequential responsive-projective relations between contributions to discourse and assumptions about relevant topics and activity types) pragmatics comes in early, and some aspects of linguistic analysis only afterwards, rather than the other way around (as was often assumed in linear models of generative psycholinguistics). Now, if SCRMs are determined in situ, and interdependent with interactional-sequential and activity-type-related aspects of situations, it would be natural to assume that linguistic resources too reflect this context-sensitivity; they ought to be designed to be used in particular types of context. All kinds of lexical items tend to be semantically polysemous and underdetermined (in relation to possible situated requirements) at the abstract lexical level, with their polysemy partially eliminated and their meanings enriched in particular situations. Many compound words have conventionalised ("lexicalised") meanings that cannot be derived from their constituent semantic morphemes. For example, a compound like Swedish *kvinnohus* or German *Frauenhaus*, both literally meaning `women´s house´, refers to a place where women can get help or shelter, or simply be provided mutual assistance. Yet, it has been reported that persons who do not know Swedish or German language (and culture) well enough, have sometimes believed that these words denote something completely different, e.g. `brothel´.

---

[6] Recanati (2004) accounts for several different kinds of semantics from context-independent ("insensitive") alternatives to more contextual or context-interdependent ones.

In Construction Grammar (e.g. Croft & Cruse, 2004), it is usually assumed that grammatical constructions, like the lexicalised compounds above, tend to have construction-specific semantic properties that cannot be derived by rule. Yet, some (Kay & Michaelis, forthc.; cf. also Croft & Cruse, op.cit.: 250ff) have raised counter-arguments on the ground that there are semantic regularities in the make-up of compounds and (some) idiomatic constructions that language users can exploit. For example, in *kvinnohus* and *Frauenhaus* one will obviously see the relations to the constituents *kvinna/Frau* \`woman´ and *hus/house* \`house´. However, Croft & Cruse (p. 250-252), following Nunberg, Sag & Wasow (1994), even argue that there is a kind of compositionality involved in an idiom like *spill the beans*, meaning \`to divulge information´, because one could map the verb *spill* (and its meaning) onto \`divulge´, another verbal unit, and similarly *the beans* onto \`information´ (noun phrases). Contrary to this argument, it can hardly be denied that one must know the expression as a whole (*kvinnohus*, *spill the beans*) and its meaning (potential) in order to produce "correct" ("idiomatic") utterances in the language in question. The fact that we can sometimes understand utterances involving non-compositionality (more or less correctly) does not demonstrate their compositionality (in a strict sense), since after all we use available contextual information in the process of understanding. What is undoubtedly there is merely some semantic regularities (or partial compositionality, if you prefer that concept).

Yet another reason for doubt in semantic compositionality is that those who insist on it tend to analyse regimented sentences, rather than practitioners´ often "illogical" ways of expressing themselves in spontaneous, interactional languaging. In fact, these latter modes of expression are often quite normal in authentic languaging, and yet they are either ignored or regarded as ungrammatical in normative written-language-biased grammars (cf. Linell, 2005, 2009, 2013). Such phenomena include incomplete utterances, composite utterances (e.g. language, manual gestures and gaze) (Enfield, 2009) and utterances involving in-course-of-progression changes of speech act status.

In short, it seems reasonable to doubt the claims of compositionality, that is, the idea that a language must be a code, also at the level of formal semantics. It seems probable that linguists´ often unquestioned belief in the need for semantic representations of whole sentences is a reflection of an assumption that language must be a code (at least in some sense) in order to function in actual communication (cf. Pagin, op.cit.). But in a reasonably strict sense it simply cannot be true. Indeed, in the light of the importance of contextual resources (by definition outside of the linguistic utterances themselves), one may even conjecture that participants in communicative practices do not need semantic representations at all; it would be sufficient with meaning potentials of lexical items and grammatical constructions in combination with the contextual factors. Compositionality of complex expressions in a strict sense might hold for exceptional types of genres and activity types, in which the exact wording of sentences, primarily in writing, is crucial. But such cases are relatively rare, and cannot be taken as basic in languaging.

Perhaps, the search for compositionality is best motivated by the need for explicit representations in computer-based applications. Of course, it *could* be the case that partial computations of *some* semantic compositional structures is part of utterance understanding more generally. Therefore, it seems worth trying out how far you can get in working with explicit semantic representations. Yet, there are still good reasons to assume that human beings, with their embodiment and socio-cultural embeddings, are not so computer-like.

Phonology

Let us now turn to phonology. Is there a "phonemic code" in languages? This is a rather different issue than the above-mentioned issues of semantics and pragmatics. Phonology is another kind of level of "linguistic articulation", one at which the elements ("phonemes") do not carry (referential, conceptual, emotional or coordinative-interactive) meanings by themselves (though prosodies may be a boundary case).

Before turning to structuralist approaches to phonology, let me just point out that a dialogist approach (Linell, 2009) to practices and processes in language and languaging would arguably lead us to look at phonetic behaviour in terms of *gestures* and actions (e.g. Fowler, 2010). Such a perspective would be less interested in static signs, sounds, phonemes, or other entities and abstract objects. Phonological gestures are:

- dynamic *gesturings* rather than static forms (expressions portrayed as stable ”signs”);
- parts of embodied actions, with goals and meanings, i.e. they are not involuntary movements;
- defined by target values[7], which are partly segment-oriented;
- parts of temporally distributed and sequentially ordered, multimodal, larger wholes, with prosodies and coarticulations; at the phonological level, such sequences are packaged as word gestalts (phonological words and phrases), i.e. they must be recognisable as words in the language;
- vary in execution/performance;

- in their combinations with prosodic gestures;
- emotional differences expressed by prosodies and voice quality variations;
- in terms of hyper/hypo-articulation; gestures need not be completely executed , in fully articulated forms;
- with regard to other sociolectal variants: “lects” contribute to speakers´ social identities and voice exploitations (speakers can imitate other people´s voices or dialects).

Phonology is centred around segments and invariants (at some level of abstraction: phonemes). Lindblom (this volume) emphasizes the combinatorial structure of sound patterns in phonologies, something which makes the amazing size of vocabularies in different languages possible. Words appear to be based on combinations of “units”. In my terminology of above, these units come in the shape of segment- and syllable-based packages of articulatory gestures (each package involving a coordination of many muscular actions: lip-, tongue- and jaw-movements, etc.). The reality of such packages seem to be proved, for example, by permutations (such as our *queer old dean* for intended *dear old queen*) and other “slips of the tongue”. Thus, segments have a natural basis in the biology of speech, yet the phonemic analysis appears to have been enhanced by alphabetic writing (see below).

As regards *variation*, it is a universal and controlled aspect of languaging. It is partially intentional, as in different speech styles, but also automatised in individual and social habits (”voices” and idiolects vs. dialects). These habits are highly functional in that they allow listeners to identify individuals and their social-group membership(s). Variation must therefore be explained as part of language and languaging. Does the same also hold for the *invariants*? How should we understand invariants? Where and what are they? Are phonemes invariants? Is an invariant an abstract entity, an exactly defined linguistic value, a point in the acoustic space, or a fuzzily defined range of variation, a cloud of points defined by their equivalence with regard to linguistic function in words? Are invariants a result of thinking in terms of linear input-output models in the psychology of perception (cf. Järvilehto et al., 2011)? Are invariants also a heritage from structuralism and the “written language bias” (Linell, 2005) of linguistics? There don´t seem to be any ready-made answers to such questions. But one point should be made; variation is also part of language, despite the attempts of many structuralists to exclude it from their models.

Basic phenomena in languaging are dynamic embodied processes from which ”basic units”, in phonology and other domains of language emerge (Lindblom, Diehl, Park & Salvi, 2011; Lindblom, this volume);

---

[7] Compare the notion of ‟motor equivalence” (Lindblom, this volume). This reflects, if you will, the ”formal” (structural) aspect of $^{\text{languaging}}$ (constraints on languaging, cf. Rączaszek-Leonardi, 2011).

they are not innate pre-specified systems of units, as generativists have proposed.[8] This emergentism might entail a refutation of classical structuralism (Lindblom, this volume). But we should equally put a question mark around *elementarism*: the idea that larger units are always built from-bottom-up (elements at one level being constituents at the next level higher up). Note that this is often presented as an aspect of structuralism too. But if, instead, we approach human activities from the phenomenological end, we conclude that they have gestalt properties of human sense-making, which calls for a partial top-down, moderately holistic approach. That is, speakers and listeners start out from some pragmatic assumptions about what their communicative projects as larger wholes are about (section "Pragmatics"). This is *partial holism*, not wholesale (abstract) holism which is what one usually finds in linguistic structuralism (Linell, 2009). On the other hand, we may perhaps argue that structuralism and self-organisation suit phonology better than syntax, lexicon and (especially) interaction norms (pragmatics).

Lindblom et al. (2011; Lindblom, this volume) have shown how general constraints on phonological systems emerge from general principles. But phonological elements are distinctive of whole words, rather than by themselves. Saying a word, e.g. Engl. *pat,* does not mean that one chooses sound by sound: /p/, then /æ/, then /t/. In addition, we cannot assume that linguistic expressions are composed only of phonemes, just as written words consist of letters and diacritics. But perhaps we could say that phonological gestalts are composed by phonemes and prosodies, i.e. that some kind of compositionality is at hand.

One area in which there has been talk about "the code" revolves around learning to read. Children who have acquired an ability to map letters and letter combinations onto sounds and words, and vice versa, are said to have "broken the code". But notice that this is only one prerequisite for reading and writing. The overall goal is that of reading for content, and writing about something. The "why" of communication (section "Pragmatics") remains to be accounted for.

Accordingly, even if we may talk about a kind of compositionality at the phonological level, this would be quite a limited kind of compositionality. We are no longer talking about expression-meaning pairings. Another point is that what is derived by general constraints (Lindblom et al., op.cit.) is not the language-specific phonological system, but a general phonology. When we move on to language-specific phonologies, we must include the sociocultural aspects. Before proceeding to the idea of genetic codes, I shall dwell for a while on cultural language.

The cultural side of language

When we talk about the "natural languages" of the world, we must not forget that these have not emerged exclusively from "natural", e.g. biological or Darwinian, processes in phylogenetic evolution. They are also cultural, especially in those cases where they have literate, written counterparts, emergent in and through a sociocultural history ("glossogenesis", Rączaszek-Leonardi, 2011). Thus, linguistic forms "reflect cultural preoccupations and ecological interests […], as well as general biological and psychological constraints […]" (Hodges, 2011: 148). Yet, despite the impact of literate society on language-specific phonetic (phonological) processes, we often assume that spoken languaging must be explained solely in natural-science-based terms. But actually, our ideas about language are penetrated by experiences of writing (Linell, 2005), as argued below.

Writing is selective; of the many phonetic properties present in speech, only relatively few are codified. Phonetician Robert Port (2006, 2011) argues that phonemes are "conceptual blends" (Fauconnier & Turner, 2003) of letters and speech sounds. The evolution of the alphabet involves selecting a few aspects of phonetic structure according to the schema: Invariance & variation in natural speech >> selection in

---

[8] As MacNeilage (2008) argues, Chomsky and generativists in general take the stance that language "just is there" and that no cogent explanations of its evolution can be produced. See also Lindblom (this volume) and Anward (this volume).

order to provide economical writing systems >> various kinds of spelling >> conventionalisation/standardisation of orthography >> repercussions on meta-awareness of language (phonology) and actual speech.

Learning to write and read has always been and is still an intense learning process both socio-historically and ontogenetically. It involves institutionalised schooling. Research into the sociohistorical genesis of literacy has shown that beginnings often involve different scribal practices (e.g. Johnson, 2003), which are later unified into national standards. Such standards can select different solutions. Relatively similar languages, such as Danish and Swedish, have chosen different ways of spelling stressed short vowels followed by (relatively) long consonants: Sw. *att*, *hatt*, *pick och pack, späck* -- Dan. *at*, *hat*, *pik og pak*, *spæk;* long consonants have double signs in written Swedish (<ck> above is the conventional orthography for /k:/ or /kk/), but not so in Danish.

Conventional orthographies are often rather far removed from being strict encodings of phonemes in speech. Hockett (1958: 541) calls English spelling a "horrible example". For example, the phonemic unit /əu/ in English can be spelled <o> (as in *go*), <ow> (as in *blow*), <oa> (as in *boat*), <ough> (as in *though*), etc. Conversely, the polygraph <ough> corresponds to several phonemic units, besides /əu/ (*though*), /u:/ (as in *through*), /ɔ/ (as in *ought*), /ɔf/ (as in *cough*), /ʌf/ (as in *tough*), and several others. And so on, in many many-to-many relations.

Once conventional orthographies have been reasonably well established, writing may have repercussions on spoken language (so-called "spelling and reading pronunciations"; e.g. Teleman, 2003: chapter 5): in Swedish, certain nominal plural endings change from [ər] in *kvinner* `women´ to [ʊr] (cf. standard orthography: *kvinnor*), a distinction between the supine form (*skriv –it*) and the perfect participle (*skriv-et*), both corresponding to English `written´, is introduced, first in writing and then in talk, and there are other examples of new "spelling pronunciations": *av* (instead of *å(v),* `of, from´), *åk* (instead of å [ɔ] *och* `and´), etc.

Another kind of evidence comes from children´s spontaneous spellings, that is, self-made spellings that children use before they have acquired standard orthography. Swedish children, for example, regularly notice the difference between the tense, aspirated, unvoiced stop consonant in e.g. *par* `pair´ and the lax, unaspirated, voiced stop in e.g. *bar* `bar´*,* a phonemic distinction in Swedish. But what about the unaspirated, voiceless variant in e.g. *spar(a)* `to save´*?* Well, it is not so seldom that we find the alternative *sbar* in children´s spellings; they have then noticed the absence of aspiration (and perhaps some laxness) in the stop, and chosen a spelling which we do not use in conventional orthography. Once, speakers have learnt to spell properly, they start to treat the labial stop consonant in *spar as if* it was simply a /p/ or [p]. Indeed, they acquire adults´ belief that it *is* a /p/, Other examples are the vocalic bursts after the stop and before the /r/ (in casual, rather than hyper-articulated, speech) that we believe are there in e.g. Swedish *opera, kamera, Monstera* but not in e.g. *Sandra, ockra, umbra, Kleopatra*.

An alphabet writing system reinforces our belief in the phonemic principle. A syllabic writing system, on the other hand, may suggest a syllabic principle in phonological theory (as in prosodic or autosegmental phonology). The matter is partly confounded, however, since actual writing systems are seldom purely segment-based. In general, the selection of phonetic affordances for phonological use in specific language systems cannot be completely predicted in an algorithmic manner.

To sum up, phonology, like other domains of language, must be approached from two, diametrically opposed ends:

- biology/physiology (a researcher´s perspective): the theory of bodily mechanisms and processes of the human organism;
- socioculture and phenomenology (partly a participant´s perspective): how we experience languaging and assign meanings in everyday reality. This features a "constructionist" aspect, for example, as regards how the social environment reacts to ("hear") the words and sounds that the

> child makes. Socioculture and phenomenology are also very much about (the theory of) the content of sense-making, which is largely about the world outside organisms.

On the other hand, how we assign weight to these different interests depends on the kinds of phonological problems we deal with. For those interested in *general* principles of phonological systems, and underlying processes of self-organisation, evolutionary biology takes the lead. If, by contrast, our research endeavours focus on *language-specific* details of phonologies, specific sociohistorically grounded cultural properties bring in the latter dimension too.

Linguistic codes?

Let me try to sum up some conclusions: Can we find code-like relations in language and languaging? To bring some order to this discussion, we had better distinguish between expression–expression relations and expression–meaning relations. I start with the former.

In the domain of natural language, are (strings of) units of expression strictly encoded into (strings of) other units of other kinds or at other levels? For example, the make-up of lexical units (morphemes, words) seems to be built (exhaustively encoded and hence decodable) in terms of combinations of units drawn from a finite set and limited set of unit types (phonemes and prosodies). But this encoding involves a considerable abstraction in that it neglects individual and social variation. Furthermore, it ignores the fact that words (and larger expressions) are associated with meaning potentials that cannot be derived from phonology or phonetics. In languaging, participants express themselves in words, grammatical constructions and contributions to communicative projects, not in terms of the constitutive phonological units.

Phonology involves phonological–phonetic gestures and phonological gestalts. That is, whole word forms, sometimes extended by means of cliticisation of function words, often make up complex phonological gestalts (cf. “phonological words”), in which the segments of the corresponding citation forms may be eliminated or only leave traces in the form of colouring (nasalisation, palatalisation etc.) of syllables and their remaining segments in presto speech or “sloppy pronunciations”. Yet, the forms can be recognised and “understood” by language-competent listeners (gestures need not be fully realised, see section “Phonology”). But the relation of phonemic structure and situated realisations in speech is not strictly code-like.

Is written representation of speech a case of recoding speech into writing? Well, there are clearly regular rules for how to spell words and for how to read text aloud, but these do not live up to requirements of a strict code. Written forms are not a code for phonemic forms, nor do letters map on to phonemes. As compared to writing, spoken language has a marginally but still significantly different encoding in terms of constituent units (phonemes and prosodies vs. letters and diacritics). In neither direction is the speech <–> writing translation strictly code-like; it is not like the mapping of letters onto Morse signals. But notice that the idea of codes becomes considerably more plausible as soon as we restrict ourselves to writing-internal relations! This is a case of something much wider, the “written language bias in linguistics” (Linell, 2005).

None of these en/de/re-coding relations involve full-fledged expression-to-meaning mappings, let alone in the other direction (meaning-to-expression). There are no code relations between complex expressions and situated meanings (SCRMs), and no code relations between meaning potentials and SCRMs, since the latter are also crucially dependent on contextual resources, the relevance of which cannot be predicted in the specific cases.

Language and genetic codes

It follows from the previous sections that we ought to be very cautious in making analogies between language and the genetic code. Possibly, the case of phonology might be a less dangerous analogy. But the notion of "genetic code" itself seems conceptually dubious. Some people talk about "reading the genetic code". Freedman (this conference) and others (Willson, 2011) are more cautious, talking about transcribing and translating, when referring to molecular biology, and, in particular, the relations between DNA, RNA and proteins.

It has been common to talk about genetics in terms of "codes" and often the analogy with natural language has been suggested (e.g. by François Jacob and Roman Jakobson) (cf. Linell, 2012a). The "genetic code" has been seen as a kind of instructions for the organism to use in governing the development of organs and their functions. The question arises what kind of model of language comes to mind in such a context? As Markoš & Faltýnek (2011) point out, it seems to be one of strings of letters to be "read" or at least decoded. As these authors go on to argue, this is a strange idea, given that we are faced with causal, metabolic processes involved in living, self-organising systems. They therefore suggest that we need a model of spoken languaging rather than written language. The discourse about "reading the genetic code" is a case of the "written language bias" (Linell, 2005), in this case in a genre of popularising scientific discourse.

A neurophysiological structure, or system of processes, cannot be "read". Who would be the reader in such a case? A homunculus sitting somewhere in the brain? The root metaphor of "reading the genetic code" has its source in the human being reading a text. But that text has a "content"; it represents something in the (real or fictive) discourse world, the topics that the text is "about". In the normal type of reading, people use marks on paper or a computer screen *to make sense about something else* than the text itself; we are not just reading off the marks (identifying categories of letters, or computing a purely "linguistic meaning"). Nor is reading just the de- and recoding of marks as in a scanning machine. (In this context one should also be aware of the conceptual difference between information and knowledge.)

"Translation" is also a metaphor, when it is used about the transduction between biochemical substances as in molecular biology (Willson, 2011), or between different material representations and procedures in digital technology. The source of the metaphor is arguably that of translation of texts between different human languages. Metaphors always build upon a selection of (often abstract) similarities between sources and targets. It is important to explore these similarities carefully; what does the metaphorical as-if relation really involve? All too often are models in science taken to be factual accounts, when actually they are at best reasonably good metaphors, analogies or as-if comparisons.

When we try to compare ”genetic codes” with ”language”, it is therefore important to determine more exactly what we mean by ”language", "translation", "code" etc. Written language is not a viable alternative. In view of the arguments raised above about, in particular, situated communicatively relevant meanings pertinent also to spoken, interactional languaging, I am also inclined to refute the attempts to use spoken language as a possible metaphor source (thus casting doubt on Markoš & Faltýnek ´s proposal too). Nor am I convinced that biosemioticians (e.g. Barbieri, 2011) have explained the putative difference between information and meaning in biology. What we perhaps can do, however, is to draw analogies between self-organisation in phonology and other self-organising systems in human biological evolution.

Conclusions

I have drawn renewed attention to the fact that so-called natural languages are interdependent with both nature (biology) and culture (sociohistory). Cultural environments, such as the use of alphabetic writing, implies that specific features of phonetic realities are selectively attended to. They also have a considerable impact on the scientific development of linguistics, including phonology.

There are strong reasons to assume that no natural language is a code, that is, a set (or several sets) of fixed expression-expression or expression-meaning pairings. This implies that analogies between natural language and the so-called genetic code need to be used with great caution. However, it may be that phonological systems are, as far as natural language is concerned, a primary case of self-organisation, which is one property associated with some codes. However, it is not the only one.

References

Anward, J. (2011). Doing Language. Ms. Linköping: Department of Culture and Communication.

Anward, J. (This volume). On the open secret of language.

Auer, P. (2005). Projection in interaction and projection in grammar. *Text*, 25: 7-36.

Auer, P. (2009). On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences*, 31. 1-13.

Barbieri, M. (2011). A mechanistic model of meaning. *Biosemiotics*, 4: 1-4.

Bilmes, J. (1985). “Why that now?”: Two kinds of conversational meaning. *Discourse Processes*, 8: 319-355.

Bruner, J. (1973). *Beyond the Information Given*. New York: Norton.

Carston, R. (2002). *Thoughts and Utterances: The pragmatics of explicit communication*. Oxford: Blackwell.

Croft, W. & Cruse, A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Enfield, N. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances.* Cambridge: Cambridge University Press.

Fauconnier, G. & Turner, M, (2003). Polysemy and conceptual blending. In Nerlich, B., Todd, Z. , Herman, V. & Clarke, D. (eds.), *Polysemy: Flexible Patterns of Meaning in Mind and Language*. Berlin: de Gruyter. 79-94.

Fischer-Jørgensen, E. (1975). *Trends in Phonological Theory.* Copenhagen: Akademisk Forlag.

Fowler, C. (2010). Embodied, embedded language use. *Ecological Psychology*, 22: 286-303.

Hockett, C. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Hodges, B. (2011). Ecological pragmatics: Values, dialogical arrays, complexity, and caring. In Cowley, S. (ed.), *Distributed Language*. Amsterdam: John Benjamins. 135-160.

Jakobson, R. (1972). Language in Relation to Other Communication Systems. In *Selected Writings*, Vol. 2. The Hague: Mouton. 570-579.

Johnson, R. (2003). *Skrivaren och språket: Skriftspråksbruk, kasus och vokaler I medeltidsbrev på svenska.* (`The Scribe and the Language: Literacy, Cases and Vowels in the Medieval Charters in Swedish´). Göteborg University: Department of Swedish.

Järvilehto, T., Nurkkala, V.-M. & Koskela, K. (2011). The role of anticipation in reading. In Cowley, S. (ed.), *Distributed Language.* Amsterdam: John Benjamins. 15-31.

Kay, P. & Michaelis, L. (Fortchc.). Constructional meaning and compositionality. In Maienborn, C., von Heusinger, K. & Portner, P. (eds.), *Semantics: An International Handbook of Natural Language Meaning*. Berlin: de Gruyter.

Kravchenko, A. (2007). Essential properties of language, or, why language is not a code. *Language Sciences*, 29: 650-671.

Lindblom, B. (This volume). The units of speech: a bio-cultural perspective.

Lindblom, B., Diehl, R, Park, S.-H. & Salvi, G. (2011). Sound systems are shaped by their users: the recombination of phonetic substance. In Clements, N. & Ridouane, R. (eds.), *Where Do Features Come From? Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*. Amsterdam: John Benjamins. 67-97.

Linell, P. (2005). *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations.* London: Routledge.

Linell, P. (2009). Rethinking Language, Mind and World Dialogically: Interactional and Contextual Theories of Human Sense-Making. Charlotte, NC: Information Age Publishing.

Linell, P. (2012a). On the Nature of Language: Formal Written-Language-Biased Linguistics vs. Dialogical Language Sciences. In Kravchenko, A.V. (ed.), *Cognitive Dynamics in Linguistic Interactions.* Newcastle upon Tyne: Cambridge Scholars Publishing. 107-124.

Linell, P. (2012b). Pragmatics and Situations First, and Language and Linguistic Analysis Afterwards. Paper presented at CILC 2012, Odense, Denmark, September 2012.

Linell, P. (2013). The Dynamics of Incrementation: Processes and resources in utterance production. Forthc. In Szczepek Read, B. & Raymond, G. (eds.), *Units of Talk – Units of Actions.* Amsterdam: John Benjamins.

Love, N. (2007). Are languages digital codes? *Language Sciences,* 29: 690-709.

MacNeilage, P. (2008). *The Origin of Speech*. Oxford: Oxford University Press.

Markoš, A. & Faltýnek, D. (2011). Language Metaphors of Life. *Biosemiotics,* 4: 171-200.

Norén, K. & Linell, P. (2007). Meaning potentials and the interaction between lexis and contexts: Some empirical substantiations. *Pragmatics*, 17: 387-416.

Nunberg, G., Sag, I.A. & Wasow, T. (1994). Idioms. *Language*, 70: 491-538.

Pagin, P. (2003). Communication and Strong Compositionality. *Journal of Philosophical Logic*, 32: 287-322.

Port, R. (2006). The graphical basis of phones and phonemes. In Munro, M. & Bohn, O-S. (eds.), *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production.* Amsterdam: John Benjamins. 349-365.

Port, R. (2011). Phones and Phonemes are Conceptual Blends, Not Cognitive Letters. Ms. Indiana University: Department of Linguistics and Cognitive Science.

Rączaszek-Leonardi, J. (2011). Symbols as constraints: the structuring role of dynamics and self-organization in natural language. In Cowley (2011): 161-184.

Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.

Teleman, U. (2003). *Tradis och funkis: Svensk språkvård och språkpolitik efter 1800.* (skrifter utgivna av Svenska språknämnden, 87). Stockholm: Norstedts Ordbok.

Willson, P. (2011). Translation as a metaphor in scientific discourse. *Translation*, inaugural issue 2011; 82-83.

Compositionality, Complexity, and Evolution

Peter Pagin

Department of Philosophy, Stockholm University

peter.pagin@philosophy.su.se

Introduction

It is widely believed that human cognition has evolved together with human language. It is also widely believed that the compositionality of natural language is one of its central features. The question then naturally arises why, and how, language evolved into a compositional means of communication.

Compositionality allows the body of semantic knowledge to be considerably smaller than the piecemeal knowledge of the language, i.e. independent knowledge of the meaning of each of the sentences of the language. Having a language with a greater expressive power seems like a good thing for a population, and maybe that could be enough for answering the question. It isn't, however.

One thing that can be asked further is by what evolutionary mechanisms a compositional language evolved. We don't believe that there was a sudden leap from an un-systematic to a compositional language, and we may ask what principle may lead, step by step, from the former to the latter[9]. In the section "The complexity of learning" I'll review an attempt to answer this question by appeal to the cognitive difficulty and computational complexity of *language learning*.

In the section "The complexity of interpretation" I shall present a different aspect of computational complexity as applied to the cognition of language, this time concerning the complexity of interpretation. I shall spell out the relation to compositionality, and end with a suggestion of how this connects with questions of evolution.

In the next section, I shall give a brief introduction to the idea of compositionality.

Of course, this is not convincing, since a small change in a rule system, or in a hard wiring, can create recursion, or feedback loops.

Compositionality

There are some remarkable facts about human linguistic communication: language users manage to successfully communicate new thought contents (contents not thought before) by means of new sentences (not used or considered before). The first part, that *new contents* are communicated, is a phenomenon that could happen simply because of context dependence: if 'kraa' means *food here now*, then every new utterance of 'kraa' means something different, even though the sophistication of speaker or hearer need

---

[9] In a public talk in Paris, 29 May 2010, entitled 'Poverty of Stimulus: Some Unfinished Business', Noam Chomsky commented on the evolution of language literature, claiming that it does not concern the evolution of language. This is because it suffers from the prejudice that evolution must take place in small steps. However, Chomsky remarked, the step from a finite to an infinite linguistic capacity is anything but small. So it cannot have been a step of evolution, as long as that is understood according to current theories.

not exceed that of seagulls. However, humans, unlike seagulls, appear to be able to communicate new contents, from an open-ended set, even if context is held fixed[10]. When context is held fixed (at least in relevant respects), if a different content is communicated, this is because a different sentence has been used[11]. So the human hearer H appears to have the ability to associate a new sentence uttered by the speaker S with a (probably new) *content* that is the *same* as, or at least sufficiently similar to, the content which S wanted to communicate, and the speaker S appears to have the ability to *select* a suitable sentence s that makes the task of H feasible. The speaker therefore has an *articulation* (production) capacity that interlocks with the *interpretation* (comprehension) capacity of the hearer. Since mind-reading is out of the question, and chance would deliver a microscopic rate of communicative success, given the sizes of the domains of contents and sentences, there must be something about language itself that makes this possible[12].

The interpretation part of this coordination task has received much more attention than articulation part. The hearer's task is made possible if there is a systematic correlation between the build-up of a sentence and its meaning. This is usually stated as the principle of *compositionality*:

(PC) The meaning of a complex expression is a function of the meanings of its parts and its mode of composition.

We shall say that the *semantics* of language is compositional or not, intending the assignment of meaning by some particular semantic function or semantic theory. In a derivative sense, we can say that *meaning property* of a particular language is compositional, if the *correct* semantics for it is compositional in the first sense.

The picture that the appeal to (PC) gives of the hearer's capacity is this: The hearer knows the meaning of the simple expressions of the language (lexical items and morphological units), which are finitely many. The hearer also knows the semantic significance of the (morpho-) syntactic modes of combination, which are finitely many. Putting knowledge of these two kinds together, the hearer can work out, step by step, the meaning of complex expressions. He starts with working out the meaning of those complex expressions that have simple parts, and then using this knowledge he can go on with combinations that have these complexes as parts, and so on. In interpreting (1a), given the syntactic structure (phrase structure) of (1b),

(1) a. John saw the dog

b. $[_S \text{John}_{NP} [_{VP} \text{saw}_{VT} [_{NP} \text{the}_{Det} \text{dog}_N]]]$

the hearer knows the meaning of 'the', 'dog', 'John', and 'saw', and the semantic significance of a) the Determiner + Noun (Det+N) combination, b) the Verb (transitive) + Noun Phrasecombination(VT+NP),and c) theNounPhrase + Verb Phrasecombination(NP+VP). He starts with working out the meaning of 'the dog', using a), goes on working out the meaning of 'saw the dog', using b), and finishes by working out the meaning of (1a) from there, using c).

Although (PC) is a standard formulation, what is meant by 'part', as the principle is normally understood, is immediate part (immediate constituent). The principle is often incorrectly stated as saying that the meaning of the complex expression is a function of the meanings of the simple parts and the way they are combined. But this principle is degenerate. The mode of combination of the simple parts in (1) is in itself a combination of Det+N, VT+NP, and NP+VP, and the speaker would have to know the semantic

---

[10] Of course, this is inferred; because of the uniqueness of contexts, it cannot be directly tested.

[11] It may also be that the same sentence has been used, but under a different syntactic analysis. More about this below.

[12] Indeed, this phenomenon has (on and off) motivated reflection on language during a hundred years. Cf. Frege 1980; Frege 1923, opening passage.

significance of that complex mode. In fact, there are denumerably many ways of combining simple parts, and so if the hearer needs to know the semantic significance of each, the explanation of how he can understand new sentences cannot any more appeal to a working out strategy based on *finite* knowledge. Moreover, on the alternative formulation, the hearer who knows that the father of Annette is also the father of Isabel is not licensed to infer that the sentences

(2) a. The father of Annette sleeps

b. The father of Isabel sleeps

have the same truth value, for he is not licensed by the appeal to only simple parts of making use of that fact that 'the father of Annette' and 'the father of Isabel' are coreferring noun phrases.

If instead of requiring that the meaning of the complex is determined by the immediate parts and *their* combination, we require that it be determined by the immediate parts and their immediate parts, and the mode of combination of these two syntactic layers, we get a principle that is strictly weaker than (PC), but not degenerate. In fact, it has been argued that some particular constructions in English obey only this weaker principle[13].

A principle closely related to (PC) is

(PS) If in a complex expression A a constituent e is replaced by a constituent *e*' that has *the same* meaning as *e*, then the meaning of the new expression *A*'= *A*[*e*'/*e*] has the same meaning as *A*.

(where '*A*[*e'*/*e*]' is designates the expression the results from substituting e' for e in one or more occurrences). The intersubstituted constituents need not be immediate. In fact, if the semantics is total, so that each grammatical expression is meaningful, then (PC) and (PS) are *equivalent*.

If, by contrast, the semantics is *partial*, the two principles can come apart. On the one hand, if not all parts of a complex expression are meaningful, then the meaning of the complex is trivially not a function of the meanings of the parts ((PC) does not hold), but it can still hold that when two parts with the same meaning (i.e. meaningful parts) are intersubstituted, the meaning of the complex is preserved ((PS) holds). On the other hand, it may also be that *A* is meaningful, and that e and e' mean the same, yet *A*[*e*'/*e*] is *not* meaningful ((PS) does not hold), while it is still true that the meaning of every complex expression that *has* meaning is function of the meaning of its parts and their mode of combination ((PC) holds).

These observations are due to Wilfrid Hodges (2001). He calls the principle that the parts of every meaningful expression are meaningful *the domain principle*. He also calls the principle that if two expressions mean the same, then substituting the one for the other does not lead to loss of meaning in the complex *the husserl principle* (and the semantics *husserlian*). He also proved that given that the domain principle and the husserl principles hold, (PC) and (PS) are again equivalent, even if the semantics is partial.

Using the equivalence, we can see that (PC) in several respects quite weak. To get a *counterexample* to (PS), and thereby to (PC), we need two expressions *e* and *e*' with the *same* meaning, and two complex expressions, *A* and *A*[*e'*/*e*] with *different* meanings. If there is no counterexample, (PS) and hence also (PC) hold. Now, if no two different expressions in the language in question have the *same* meaning (meaning is *hyperdistinct*), then no counterexample is possible, and hence the semantics is vacuously compositional. Similarly, if no two expressions in the language *differ* in meaning, again the semantics is trivially compositional.

The first of these two observations, that hyperdistinctness entails compositionality, is somewhat damaging for the explanatory power of compositionality. Meaning for complex expressions of a language may obey

---

[13] According to Peters and Westerståhl (2006, chapter 6), this holds for the semantics of possessive constructions of the form NP's.

no regularity whatsoever, so that it is impossible to *work out* the meaning of a new complex expression, and yet because of hyperdistinctness, its meaning is compositional. So compositionality alone does not explain how speakers can work out the meaning of new complex expressions.

To explain the ability of hearers, the semantics therefore needs some additional properties. In particular, the semantic function that assigns meanings to expressions must be *computable*. This raises the following problem: If the semantic is computable, why should it also be compositional? This is a problem, for, as we shall see in "The complexity of interpretation", a language may have a computable semantics that is not compositional.

So far we have been talking as if semantic values or properties are assigned to *expressions*, where expressions are understood as *surface* forms, i.e. types of spoken utterance, types of written inscriptions, or types of other concrete marks or events. Surface forms are, however, often syntactically ambiguous. Two different words may have the same surface form (homonymy), and complex expressions may on top be ambiguous with respect to scope relations (witness Quine's 'pretty little girls' camp'). Since different disambiguations often yield different meanings, we must define the semantic function for *disambiguated expressions*, such as analysis trees, or *grammatical terms*. For setting out compositionality formally, we therefore need on the one hand a syntactic framework, which specifies disambiguated expressions of the language, and the semantic function and properties defined over these. We will also need a domain of semantic values and domains of other entities that can enter as arguments to the semantic function. In the modern tradition, there are two approaches to this. In the tradition from Montague[14], the syntax is built up from categories of expressions and corresponding sets of expressions that are of these categories, and the syntactic rules contain specifications of which categories their arguments and values have. The syntax then has the form of an *algebra*, where the grammatical rules are implemented as operations from *n*-tuples of expressions (of the right categories) to expressions. Correspondingly, the semantic values are organized into types of entities, and there is a *meaning* algebra, where the entities are possible semantic values, and operations in this algebra map *n*-tuples of entities (of the right types) on entities. There is a mapping between the syntactic categories and the meaning types. The semantic function, finally, is a *homomorphism* from the syntactic algebra to the meaning algebra. The function must respect the mapping between categories and types[15].

In the more recent tradition from Hodges (2001), there is no appeal to categories and types. Instead, syntactic operations are taken to be *partial*; they are defined only for certain arguments (combinations of arguments). There is a syntactic algebra, but there is no assumption that there is a meaning algebra, only a domain of meanings where no prior structure is required[16]. We shall here follow Hodges, with some modifications. The syntax for a language L is a triple (G_L, E_L, V_L), where G_L is a *grammatical term algebra*, EL is the set of *expressions* of L, and V_L is a mapping from grammatical terms to expressions (for convenience I shall henceforth drop the subscript). G itself is a triple (T, Σ, A), where T is the set of *grammatical terms*, Σ the (finite) set of *operations* that map n-tuples of grammatical terms on grammatical terms, and A is a (finite) set of *atomic* grammatical terms. T is the closure of A under Σ, i.e. the set of terms that are generated from A by means of (possibly repeated) applications of the operations in Σ.

---

[14] Cf. Montague 1970; Montague 1973, Janssen 1986; Janssen 1997, Hendriks 2001.

[15] 7The rest of this section is somewhat technical and may be skipped by readers who only want the main ideas.

[16] 8If we require that the semantics be *recursive*, then an inductive structure in the meaning domain is needed. Cf. section on "The complexity of interpretation".

To exemplify, let A be the set {‘John’,‘saw’,‘the’,‘dog’} and Σ the set $\{\alpha, \beta, \gamma\}$, corresponding to the rules for forming Det+NP, VT+NP, and NP+VP, respectively, of example (1). Then the grammatical term that corresponds to the phrase structure analysis (1b) is

(3) γ(‘John’, β(‘saw’, α(‘the’,‘dog’)))[17].

The V function compositionally defined on the grammatical terms. For each operation σ on terms, there is a corresponding operation σ on expressions such that

(V) $V(\sigma(t_1,..., t_m)) = \sigma(V(t_1),...,V(t_n)$

In the very simple case of our example, we just have

(4) $V(\sigma(t_1, t_2)) = V(t_1)\text{␣}V(t_2)$

where σ is α or β or γ, and ‘␣’ marks a word boundary (space). We therefore get, after three applications of (V),

(5) $V(\gamma$(‘John’, β(‘saw’, α(‘the’,‘dog’)))) = ‘John’␣‘saw’␣‘the’␣‘dog’

which is a structural-descriptive name of ‘John saw the dog’. No abstract limitations are imposed on V (such as that the value of the first argument should appear in the expression (surface string) to the left of the value of the second argument), although a number of restrictions will be empirically motivated, in order that the hearer be able to parse the string. The algebraic framework as such is very general; it can be adapted to phrase structure rules as well as to transformations or movements.

A *semantic* function μ is also defined on the grammatical terms, mapping terms on a domain M of semantic values. Given the algebraic framework we can now state a more formal version of (PC):

(PC') For every n-ary syntactic operation $\sigma \in \Sigma$ there is a function $r_\sigma: M^n \longrightarrow M$ such that for all grammatical terms $t_1,..., t_n$ such that $\alpha(t_1,..., t_n)$ is defined and μ meaningful, it holds that $\mu(\alpha(t_1,..., t_n)) = r_\sigma(\mu(t_1),..., \mu(t_n))$.

We call the function r_σ a *meaning operation*. If (PC') holds for a semantic function μ, we say that μ is compositional. Here μ is of course left unspecified. Compositionality is a general formal property that any particular semantic function, *given* a syntax, either has or lacks[18]. A statement that a semantics is compositional, in this standard sense, therefore is a $\forall\exists^1\forall$ statement, where the existential quantifier is second-order.

Historically, the first precise statement of (PS) was given by Gottlob Frege (1892), for *Bedeutung*. Clarity about (PC) developed slowly after that, and did not reach maturity until in the mid-late 1970s[19].

The complexity of learning

The question why evolution would tend to produce language with compositional semantics can be seen from a learning point of view. Learning, in turn, can be viewed both from a genetic and from a cultural perspective. The issues involved are surveyed in Smith and Kirby 2012. In the cultural perspective, the issue is that of *iterated learning*: A speaker B learns from a speaker A, a speaker C from speaker B, and

---

[17] This is an illustration. It might be preferable e.g. to have as top operation one that combines the term that corresponds to ‘the dog’i with the term that corresponds to ‘John saw i ’, analogous to a Montagovian quantification rule.

[18] It is *another* matter that language theorists may want to adjust their *syntactic* theory for the sake of making their semantics compositional.

[19] For more on compositionality, see Pagin and Westerståhl 2010b; Pagin and Westerståhl 2010c; Pagin and Westerståhl 2010a, Janssen 1997, Szabó 2008.

so on. Without considering genetic change, we can inquire into the conditions under which such iterated learning tends to produce language change, in each step, that drifts towards a compositional semantics.

Here I shall focus on a particular proposal in Brighton 2005. Brighton, among others, emphasizes the role of the so-called *learning bottleneck*. By this is meant the fact that only a meagre subset of the whole language is actually transmitted to the learner, and the learner needs to develop a hypothesis about the underlying regularity in order to generalize to new cases. This of course is a counterpart to Chomsky's *poverty of the stimulus* argument. See e.g. Chomsky 1971. The point of that argument was that the sample of utterances the child gets is very small, and if there are no built-in restrictions on how to project the sample on to a grammar there are just too many possibilities compatible with the sample for a projection to take place at all. Hence, there must be built-in restriction, an innate universal grammar.

Brighton, and those working in the same tradition, reject this argument. In contrast to Chomsky, Brighton thinks that we can lay down general principles for learning that will generate the desired result, without the need to appeal any particular innate linguistic structures. As regards linguistic meaning, and in particular the emergence of compositional semantics, Brighton suggests how.

Brighton operates with simple abstract languages, strings over an alphabet (as is standard in formal language theory). As meanings he uses a feature-value space: the meaning of a meaningful string is a vector, consisting of one value for each feature (e.g. (2,2,2) in a space of three features each of which can take two values). The interpreted language is a set of pairs of strings and meanings. Such a language can be specified by a comprehensive list of string-meaning pairs, and the question is whether it can be specified in some more efficient way.

As the theory of automata and formal languages were developed in the 1950s, grammars that generate languages by means of production rules correlate with automata that *accept* languages. Typically, this concerned grammar. In the case of semantics, Brighton employs the idea of a special kind of finite-state machine, a so-called *transducer*. Basically, a transducer accepts certain string-meaning pairs and rejects others. For each such pair in the language the machine accepts there is a path through *states* of the machine, starting with the initial state and ending in the final, accepting, state. For each new letter of a meaningful string, as read from left to right, there is a possible branch along which the machine can move into a new state, one branch for every meaning compatible with the initial segment of the string that has been considered so far. After the final letter, the machine moves into the accepting state, and if the string happens not to have a meaning in the language, the machine just halts in a non-accepting state.

In the most elementary, non-efficient case, there is a 1−1 correlation between strings of the language and paths of the machine. But for some languages it is possible to *compress* the machine. For instance, if two strings start with the same letter and also are associated with the same initial feature-value, the initial segments of the two paths can be *merged* into a single *shared* segment of both paths. And in general, when a letter value part is shared between two string-meaning pairs, there is a corresponding merge to be made in the path. The machine gets compressed. The *description* of the machine becomes shorter.

In case the semantics of such a language is compositional, each letter of the alphabet corresponds to a particular value of some feature. In such a case, compression of the machine can go very far, resulting in a minimal transducer with one state for each string position / semantic feature, and between these states one branch (edge) for each value of the corresponding feature. Such a machine is minimal. In the opposite case, where no compression, or hardly any compression, is possible, the language is called "random" or "holistic"; such a language has to be learned piecemeal, one word at a time[20].

---

[20] This is not the customary use of the term 'holistic' in the philosophy of language. There, it typically indicates interdependence in meanings between different expressions, and therefore in this sense semantic holism is the opposite of semantic randomness. Cf Pagin 2006.

We can therefore see that there is a certain connection between compositionality and minimal *complexity*. Complexity can be measured with respect to different parameters and in different ways. Brighton appeals to *Kolmogorov* complexity (or *minimal description length complexity*) (cf. Li and Vitányi 1997): given some standard encoding (such as a particular universal Turing machine), complexity is measured in the number of bits required to describe certain data or functions. For instance, if, given a way of encoding Turing machines by means of some particular universal machine, a smaller Turing machine can be used for computing a function *f* than can be used for computing a function *g*, function *f* is deemed to have *lower* Kolmogorov complexity than *g*.

Brighton applies to a language learning situation where the learner gets certain linguistic data (a sample of string-meaning pairs) and forms a hypothesis about the data (in the form of a transducer). The complexity is then measured as a certain sum: the sum of the length of the description of the machine and the length of the description of the data by utilizing the machine. If the machine is compact, the complexity is low. Therefore, roughly, low complexity corresponds to compositionality.

This is then implemented in an evolutionary framework, i.e. here in iterated learning framework. The learning bottleneck has the effect that in each transition from teacher to learner, only a proper subset of the entire language can be actually exhibited, and the learner needs to form a hypothesis about the language that fits the data. Brighton studied what happens when it is *built in* to the evolutionary process that learners try to minimize complexity, i.e. try to find the most compact machine compatible with the data, and *also* generate *new* string-meaning pairs only if they are compatible with the hypothesis they have formed. It was shown in computer simulations that a process starting from an arbitrary language, after a number of learning generations, eventually leads to a stable state where the language approximately compositional.

This is an interesting result. Still, from the point of view of my own interests, it suffers from two interrelated drawbacks. On the one hand, compositionality is built into the evolutionary process in the sense that it governs the *hypothesis formation* of new learners, as well as their creative production of new string-meaning pairs (with random creative production, no compression results). That limits the value of the explanation. Especially, building in a compositional hypothesis formation function does not offer a very distinct alternative to Chomsky's innateness hypothesis.

On the other hand, compositionality is contrasted only with randomness ("holism"). No account is taken of semantics that can be given a compact description by means of some *recursive* semantic function that is still not compositional. This may be a theoretical oversight, but it is also connected with the simplified semantic model in terms of acceptability by transducers. Compression depends on the association of letters with values, and if two letters are associated with the same value, they will be interchangeable, in the strings of a fully compressed language, without changing the acceptability of the string-meaning pair. By contrast, if we were to use a Turing machine for accepting string-meaning pairs, a compact machine could accept a language with infinite string-meaning pairs but a non-compositional semantics. Then the question arises again: why would compositional semantics be desirable, over and above a computable semantics? I shall propose an answer to this question in the next section.

The complexity of interpretation

The function version of compositional semantics is given by recursion over syntax, but that does not imply that the meaning operations are defined by recursion over *meaning*, in which case we have *recursive semantics*. Standard semantic theories are typically both recursive and compositional, but the two notions are mutually independent. For a semantic function μ to be given by recursion it must hold that:

Rec(μ) There is a function *b* and for every $\alpha \in \Sigma$ an operation $r\alpha$ such that for every meaningful expression *s*,

$$\mu(s) = \begin{cases} b(s) & if\ s\ is\ atomic \\ r\alpha(\mu(u1), \ldots, \mu(un), u1, \ldots, un\,) & if\ s\ =\ \alpha(u1, \ldots, un\,) \end{cases}$$

For μ to be recursive, the basic function *b* and the meaning composition operation $r\alpha$ must themselves be recursive, but this is not required in the function version of compositionality. In the other direction, the presence of the terms $u1$,..., un themselves as arguments to $r\alpha$, has the effect that the compositional substitution laws need not hold. Substituting a subterm for another subterm with the same meaning may change the meaning of the mother term[21].

By a generalization of *Church's Thesis*, a semantics is computable if and only if it is recursive. Therefore, if we want semantics to be computable, we want it to be recursive. Again, is there any reason why it should be compositional as well?

One possible reason has to do with the efficiency, and thereby with the *complexity*, of interpretation. We can look at the task of interpretation, i.e. semantic processing, as a problem in the sense of complexity theory. Classical computational complexity theory (of which the theory of Kolmogorov complexity is one species) is concerned with giving mathematical measures of the difficulty of mathematical problems. The problem need not be a problem within any standard branch of mathematics, such as number theory or geometry, but must be a problem that can be adequately represented in a formal language, as an input to computation.

For measuring the complexity of a problem one needs a computation method. One then asks how much of resources is needed by this method for arriving at a solution to the problem. A standard method that is used as a reference device in this sense is that of one-tape Turing machines. One of the standard resources is time, in the sense of the number of computation steps needed by the Turing machine for arriving at the solution. This is so-called *time complexity*.

With a problem type and computation method is associated a *time complexity function C*. This is a function that takes as argument a measure of the size of a problem instance, as a numerical value, and gives as value the size of the *largest* computation that is needed to compute any problem of the same size. We can illustrate this with one of the most classical examples, the problem of the *Travelling Salesman*: a salesman is to visit a number k of cities exactly once and then return home, and the problem is to find a visiting order that minimizes the total distance travelled. In this case the solution consists in selecting the optimal order and verifying that it is optimal. The number of cities *k* is the *size* of the problem instance, and this is the argument to the complexity function *C*. Its value *C* (*k*) is the number of computation steps needed *at most* for determining the solution for any problem instance of size *k*.

---

[21] This can happen e.g. with simple semantics for quotation, as noted e.g. in Werning 2005. Such a semantics is given by Christopher Potts (2007), incorrectly claiming that it is compositional. Cf. Pagin and Westerståhl 2010d

In complexity theory one is interested not so much in the value of C for a particular argument, but rather in how fast the value $C(k)$ *grows* when the argument increases. If $C(k)$ is bounded by a linear function of k, the time complexity is said to be linear; if it is bounded by k^n, for some natural number n, the time complexity is said to be *polynomial*, or equivalently that the problem is solvable in polynomial time.

Problems that are solvable in polynomial time are generally regarded as *tractable*, or *feasible*, while if the value of the complexity function grows faster, they are said to be *intractable* (this is known as the *Cobham-Edmonds thesis*). It is not known whether the travelling salesman problem is intractable in this sense. The reason is that no method is known for determining the solution (with certainty and for any finite k) that is more efficient than calculating the total travelling distance for each visiting order and selecting the shortest. Since the number of visiting orders for k cities is $k^{!}$, the factorial of k, and since $k^{!}$ grows faster than k^n, for any n, the general problem is intractable if there is no method sufficiently faster then checking all possible orders of travelling[22].

How does this apply to semantic interpretation? We need a method of computing meanings from disambiguated expressions as inputs. If we then think of semantics in functional terms, we want to compute a semantic function μ that takes as arguments disambiguated expressions — *grammatical terms* — and gives as values *meanings m* of some sort in this format:

$\mu(t) = m$.

Since meanings are non-syntactic abstract entities, they must be syntactically represented, i.e. by means of a sufficiently formal meta-language *ML*. That means that in an equation instance of this format, '*t* ' is replaced by an expression denoting a grammatical term, and '*m*' by an expression of *ML*.

Then we need an algorithmic method of some kind for computing meanings. A type of method that is particularly well suited is that of *term rewriting systems*. In general, a term rewriting system (a TRS) R is a pair (F, R) of a signature F and a set R of rewrite rules over that signature. The signature consists of a set of basic terms, and a set of operators. To this is added a set of *rewrite variables* which are used in stating the rules. A rewrite rule has the form

$$F(\vec{x}) \rightarrow G(\vec{y})$$

(where the arrows over the variables indicate that it is a sequence of variables)[23]. An example would be

$h(x_1)bx2 \longrightarrow g(x_1, c)bd$

where '*b*', '*c*' and '*d*' are constants. Every rule application is a substitution operation, where an instance of the left-hand-side (lhs) of the rule is replaced by the corresponding instance of the right-hand-side (rhs) of the same rule. The substitution may be per' formed on a subterm of a larger term. An instance of a term s is any term s resulting from s by uniform substitution by terms for rewrite variables. Thus, '$h(s7)bf(s9)$' is an instance of the lhs above.

A derivation is a sequence of rule applications, where every step except the initial one is an application to a term that results from a previous step. In case a term is reached such that no rule of the TRS applies to it (and hence not to any of its subterms either), the derivation has *terminated*, and the term is said to be in *normal form*. The original term is then *reduced* to normal form. A rewrite system R terminates iff every derivation eventually leads to a term in normal form. R is said to be *confluent* iff it holds for any distinct terms $s1$, $s2$, $s3$ such that $s2$ and $s3$ both can be derived from $s1$, that there is a term $s4$ such that $s4$ can be derived from both $s2$ and $s3$. R is *convergent* iff R both terminates and is confluent.

---

[22] For interesting partial results concerning this problem, cf. the Wikipedia article http://en. wikipedia.org/wiki/Travelling_salesman_problem. The problem is known to be *NP hard*, which entails that if NP =P, as is generally believed, it is intractable.

[23] For an excellent introduction to term rewriting, see Baader and Nipkow 1998.

Rewriting systems are general computation devices, in the sense that the reduction of a rewrite term to normal form is a computation. It is a standard result that any Turing machine can be simulated by a term rewrite system (cf. Baader and Nipkow 1998, , 9497). We also get a very natural measure of time complexity by just counting the number of rule applications, i.e. reduction steps, until normal form is reached. One reason why term rewriting is a natural choice for semantic interpretation is that the clauses by means of which a semantic system is defined correspond closely to rewrite rules, and can be transformed into rules by a minimal change.

To illustrate, consider Davidson's *Annette* example (Davidson 1967, 17-18) of a compositional semantics:

(6) i) Ref('Annette') = Annette

ii) Ref('the father of' $\frown t$) = the father of Ref(t)

This simple definition has the form of a system of equations, and provides a method for deriving the interpretation of 'the father of the father of the father of Annette' in four steps of substitution. Let '*F'* be the object language father operator and '**F**' its analogue in the meta-language, and let '*a*' be the object language name of Annette.

Then we have in four steps with the semantic function μa:

(7) $\mu a\ (F\ (F\ (F\ (a)))) = \mathbf{F}(\mu a\ (F\ (F\ (a)))) = \mathbf{F}(\mathbf{F}(\mu a\ (F\ (a))))$

$= \mathbf{F}(\mathbf{F}(\mathbf{F}(\mu a\ (a))))$

$= \mathbf{F}(\mathbf{F}(\mathbf{F}(\text{Annette})))$

where (what corresponds to) the second clause of (6) is applied three times and the first clause once.

Each derivation step in (7) is a substitution step. Each substitution is performed in accordance with (what corresponds to) equations in (6). These equations are applied only for substitution from left to right: an instance of the left-hand side is replaced by the corresponding instance of the right-hand side. We have then in fact used the system as a rewrite system. To make that explicit, replace the identity signs with left-right arrows:

(8) i) $\mu a\ (a) \rightarrow$ Annette

ii) $\mu a\ (F\ (x)) \rightarrow \mathbf{F}(\mu a\ (x))$

In rewrite system (8), any term of the system is reduced to normal form in a number of steps that is identical to number of symbol occurrences (i.e. occurrences of '*F*' and '*a*') of the term. If we take the size of the problem to be the size of the input term then the associated time complexity function $C(8)$ is the identity function. That is, $C(8)(k) = k$.

We c an easily speed up the system by adding a third rule:

(9) i) $\mu\mathrm{a}\ (\mathrm{a}) \rightarrow$ Annette

ii) $\mu\mathrm{a}\ (\mathrm{F}\ (\mathrm{x})) \rightarrow \mathbf{F}(\mu\mathrm{a}(\mathrm{x}))$

iii) $\mu\mathrm{a}\ (\mathrm{F}\ (\mathrm{F}\ (\mathrm{x}))) \rightarrow \mathbf{F}(\mathbf{F}(\mu\mathrm{a}\ (\mathrm{x})))$

Because of the third rule, two occurrences of '*F*' can be processed in one step. So with this addition we get another complexity function: $C(9)(k) = k/2+1$ for odd k (i.e. even number of F's), and $k + 1/2$ for even k. Clearly, by applying this method, for each system we can find another that is more efficient with respect to time. Still there is an upper bound the speed-up. Since for any system there is finite number n such that no rule application processes more than n symbol occurrences, for that system each full reduction to normal form will take at least k/n steps. Hence, no system has reductions faster than linear time.

The speed-up between systems (8) and (9) is acquired at the cost of enlarging the rule system, adding a redundant rule. Hence, we can see that there is a trade-off between the size of the system, with respect to the number of rules, and the speed of the system. It is natural to ask for the speed of a system that has a minimal number of rules, i.e. a system R such that for any equivalent system $R^{'}$, one that reduces the same input terms to the same normal form terms, $R^{'}$ has at least the same size as R. It is natural to set identity as the maximum of efficiency for such a system. That is, if for a minimal rule system R the corresponding time complexity function C_R is such that $C_R(k) \leq k$, then we say that R has maximal efficiency. $C(8)$ is maximally efficient in this sense.

For languages without variable binding, as is shown in Pagin 2012a, a semantics with minimal complexity is compositional. The conversation does not hold: just as compositional semantics can fail to be computable at all, it can be computable but not efficient. But the requirement of maximal efficiency *induces* compositionality. Moreover, if we add first-order quantifiers, complexity will still be very low, although not minimal in the sense defined (cf. Pagin 2012b). Even more significantly, if do allow recursive semantics, then the time complexity function will be *exponential*, and hence the semantics will be *intractable* (cf. Pagin 2011). And if, by contrast, we disallow recursion over meanings, the resulting semantics is compositional. Hence, the requirement even of *tractable complexity* forces compositionality.

What is the bearing of these results on language acquisition and evolution? The simple immediate ideas are these: if a language is easier to interpret, then it is also easier to learn, and language evolution would tend to favour a semantics that makes interpretation easy. But it is possible to say something more interesting.

If complexity grows rapidly, then sentences of moderate syntactic complexity can be quite hard to process. A sentence like

(10) Whenever you go into a cave, bring a torch!

is of moderate syntactic complexity, and could obviously convey useful information. With a recursive semantics, the interpretation would require very many steps, rendering the sentence unusable in practice, even though interpretable in principle. Hence, we can see why there would be a selection pressure for low interpretation complexity. Compositional semantics allows greater expressive power, simply because a systematic but non-compositional semantics of a language with the same expressive power would be intractable.

References

Baader, Franz and Tobias Nipkow (1998). *Term Rewriting and All That.* Cambridge: Cambridge University Press.

Brighton, Henry J. (2005). 'Linguistic Evolution, and Induction by Minimum Description Length'. In: *The Compositionality of Concepts and Meanings: Applications to Linguistics, Psychology and Neuroscience.* Ed. by Markus Werning, Edouard Machery, and Gerhard Schurz. Vol. 2. Frankfurt/Lancaster: Ontos, pp. 13–39.

Chomsky, Noam (1971). 'Topics in the Theory of Generative Grammar'. In: *Philosophy of Language.* Ed. by John Searle. Oxford: Oxford University Press.

Davidson, Donald (1967). 'Truth and Meaning'. In: *Synthese* 17, pp. 304–23. Reprinted in Davidson 1984, 17-36.

— (1984). *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press. Feigl, Herbert and Wilfrid Sellars, eds. (1949). *Readings in Philosophical Analysis.* New York: Appleton-Century Croft.

Frege, Gottlob (1892). 'Über Sinn und Bedeutung'. In: *Zeitschrift für Philosophie und Philosophische Kritik* 100, 22–50. Translated by Herbert Feigl as 'On sense and nominatum', in Feigl and Sellars 1949, 85–102.

— (1923). 'Gedankengefüge'. In: *Beiträge zur Philosophie des Deutschen Idealismus.* 3, pp. 36–51. Reprinted in Frege, *Logische Untersuchungen,* Vandenhoeck & Ruprecht, Göttingen, 1976. Translated by R. Stoothoff as 'Compound thoughts', published in *Mind* 72 (1963), 1–17.

— (1980). ‘Letter to Jordain 1914’. In: *Philosophical and Mathematical Correspondence.* Ed. by Gottfried et. al Gabriel. Chicago, Ill.: Chicago University Press, pp. 78–80.

Hendriks, Herman (2001). ‘Compositionality and Model-Theoretic Interpretation’. In: *Journal of Logic, Language and Information* 10, pp. 29–48.

Hodges, Wilfrid (2001). ‘Formal Features of Compositionality’. In: J*ournal of Logic, Language and Information* 10, pp. 7–28.

Janssen, T.M.V. (1986). *Foundations and Applications of Montague Grammar. Part 1, Philosophy, Framework, Computer Science.* CWI Tract 19. Amsterdam: Centre for Mathematics and Computer Science.

— (1997). ‘Compositionality’. In: *Handbook of Logic and Language.* Ed. by Johan van Benthem and Alice ter Meulen. Amsterdam: Elsevier, pp. 417–73.

Li, Ming and Vitányi (1997). *An Introduction to Kolmogorov Complexity and its Applications.* 2nd ed. New York: Springer.

Montague, Richard (1970). ‘Universal Grammar’. In: *Theoria* 36, pp. 373–98.

— (1973). ‘The Proper Treatment of Quantification in Ordinary English’. In: *Approaches to Natural Language.* Ed. by Jaakko Hintikka, J. Moravcsik, and P. Suppes. Dordrecht: Reidel.

Pagin, Peter (2006). ‘Meaning holism’. In: *Handbook of Philosophy of Language.* Ed. by Ernie Lepore and Barry Smith. Oxford: Oxford University Press.

— (2011). ‘Compositionality, Computability, and Complexity’. Draft.

— (2012a). ‘Communication and the Complexity of Semantics’. In: *The Oxford Handbook of Compositionality.* Ed. by Wolfram Hinzen, Edouard Machery, and Markus Werning. Oxford University Press, pp. 510–29.

— (2012b). ‘Truth-theories, Competence, and Semantic Computation’. In: *Davidson’s Philosophy: A Reappraisal.* Ed. by Gerhard Preyer. Oxford: Oxford University Press.

Pagin, Peter and Dag Westerståhl (2010a). ‘Compositionality’. In: *Semantics. An International Handbook of Natural Language Meaning.* Ed. by Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Elsevier.

— (2010b). ‘Compositionality I: Definitions and Variants’. In: *Philosophy Compass* 5, pp. 250–64.

— (2010c). ‘Compositionality II: Arguments and Problems’. In: *Philosophy Compass* 5, pp. 265–82.

— (2010d). ‘Pure Quotation and General Compositionality’. In: *Linguistics and Philosophy* 33, pp. 381–415.

Peters, Stanley and Dag Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Oxford University Press.

Potts, Christopher (2007). ‘The Dimensions of Quotation’. In: *Direct Compositionality.* Ed. by Chris Barker and Pauline Jacobson. Oxford: Oxford University Press, pp. 405– 31.

Smith, Kenny and Simon Kirby (2012). ‘Compositionality and Linguistic Evolution’. In: *The Oxford Handbook of Compositionality.* Ed. by Wolfram Hinzen, Edouard Machery, and Markus Werning. Oxford: Oxford University Press, pp. 493–509.

Szabó, Zoltán (2008). ‘Compositionality’. In: *Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Stanford University. URL: http://plato.stanford.edu/archives/win2008/entries/compositionality/.

Werning, Markus (2005). ‘Right and Wrong Reasons for Compositionality’. In: *The Compositionality of Concepts and Meanings: Foundational Issues.* Ed. by Markus Werning, Edouard Machery, and Gerhard Schurz. Frankfurt/Lancaster: Ontos, pp. 285–309.

Werning, Markus, Edouard Machery, and Gerhard Schurz, eds. (2005). *The Compositionality of Concepts and Meanings: Foundational Issues.* Frankfurt/Lancaster: Ontos, pp. 23–58.

Language Evolution in Humans and Ancient Microbes: What can human language acquisition tell us about the origin of genetic information?

Stephen J. Freeland and Melissa Ilardo
University of Hawaii, NASA Astrobiology Institute
freeland@ifa.hawaii.edu

Abstract

This paper seeks to encourage dialog on a question with a deceptively simple surface. When we find linguistics and genetics using the same vocabulary to describe their central phenomena, is this because the phenomena are meaningfully similar? Are we encountering a superficial analogy, enjoying the benefits of a good metaphor or recognizing some deeper principles of information organization and transfer? We approach this broad topic by focusing attention on the ancient evolutionary events that created a system of genetic coding, soon after the origin of life on our planet. Specifically we examine a progression of three topics: whether genetic code-words are arbitrary signifiers for the objects they encode (amino acids); how evolutionary biologists have deduced clues about the evolution of genetic coding by studying the complex end-product; and the current scientific paradigm for the origin of genetic information. Our suggested points of connection suggest encouraging insights from each side (linguistics to evolutionary biochemistry and vice versa), though our primary aim is to ask for further help exploring how linguistics can reshape thinking within evolutionary biology.

Keywords: evolution; genetic code; amino acid; alphabet; translation; metaphor

Background

“Whether we realize it or not, translation is the hidden lens through which almost all of human knowledge is processed. Pick any subject – religion, philosophy, science literature, history – and somewhere at its heart is a foundational work of translation”.

With this claim, Filkins (2012) summarizes an argument that a proper understanding of translation underpins academic scholarship. The argument’s literal point is easily illustrated with reference to a few well-known examples. The rational philosophies of ancient Greece passed through the Golden Age of Islam en route to shape the renaissance of Western Europe, and the Hebrew text of Genesis has flowed through numerous other languages before converging to colour contemporary political debate within the USA. In each case, an original text has undergone a process of translation in order to influence cultures widely separated in space and time. As the text moves between languages, the act of translation brings an inevitable shift in semantic content simply because each language exhibits a unique association between concepts and their corresponding expression (see, for example, Armstrong 1993 for a description of the ways in which culture, particularly religion, influenced translation of Greek philosophy; Munger, 1999 provides an excellent introduction to the challenges posed by biblical translation). Indeed classical linguistic theory teaches that even within a single language, an imperfect mapping of words to concepts forces us to undertake moment-to-moment acts of translation as we struggle to communicate our thoughts (e.g. Saussere, 1983). It is the last of these insights that forms a start-point as we now seek to travel beyond the literal truths of Filkins' quote in order to explore whether something similar holds true for the seemingly distant research domain of ancient biochemistry.

The idea of translation as a subtle modifier of meaning is the foundation from which we build, writing as evolutionary biologists who research events that occurred close to life's origins. Specifically, we are interested in the origin of genetic coding – also known as gene translation. We offer an account of what is known (and what remains unknown) about the evolution of genetic coding, with a focus on three specific features: (i) the relationship between genetic code "words" (codons) and their "meanings"; (ii) the extent to which key features of genetic coding have been optimized by natural selection; and (iii) explanations for the origin of such a system. Our intention is to solicit input from linguistics for new and fruitful ways to understand the language-like properties of the biochemical system invented by early biological evolution. With that in mind, we aim to introduce the relevant molecular biology in a manner that is accessible to an audience of non-experts, and ask you to keep in mind that it is precisely our inexperience in the field of linguistics that leads us to seek input.

We finish this introduction by noting that we consider ourselves astrobiologists – that is, participants in a multi-disciplinary scientific community that seeks to place its specialized research findings into a broader context of addressing questions of life's origin(s), distribution and future in the universe(see: Des Marais, Nuth, Allamandola, Boss, Farmer, Hoehler, Jakosky, Meadows, Pohorille, Runnegar & Spormann 2008). Beneath our explicit research question lies a motivation to clarify the roles of chance versus predictability in the evolution of life on our planet. One might infer that the existence conscious beings capable of asking and communicating such questions is the sum product of countless lucky but arbitrary events rippling into ever more unpredictable consequences through time and space (Gould, 1989). An opposite conclusion is that our species' existence is a predictable outcome, given the basic chemistry and interactions inherent in the universe (Morris, 2003). The truth seems likely to lie somewhere between these two extremes, and the scientific challenge is to find where, exactly. More accurately, we perceive the challenge is to find creative ways in which we can begin to refine estimates (see, for example, Freeland 2007). This is the theme to which we return in our concluding comments.

An introduction to biochemical language

Contemporary science uses simple linguistic vocabulary to describe the interface between biology and chemistry. All living systems "read" and "write" genetic information using appropriate chemical "alphabets". Above all, DNA genes are "transcribed" into RNA and then "translated" into protein, an entirely different chemical "language." Proteins are the metabolising entities whose functions define life as we know it, including the ability to copy genetic information into a new generation (see Figure 1).

The discoveries contributing to this "*central dogma of molecular biology*"(Crick, 1958) marked a mid-twentieth century turning point for biological science, and were rewarded with an appropriate shower of Nobel prizes (including one each for the localization of biological heredity within DNA(1969, Physiology or Medicine), the structure of DNA(1962, Physiology or Medicine), the process of genetic coding(1968, Physiology or Medicine) and that of protein folding(1972, Chemistry)). Collectively these achievements have made possible the biotechnologies of gene sequencing (Maxam & Gilbert, 1977) and genetic engineering(Hughes 2001). They have enabled scientists to reconstruct the evolutionary tree of life back to the Last Universal Common Ancestor (LUCA) of all today's living species (Williams, Fournier, Lapierre, Swithers, Green, Andam & Gogarten, 2011).

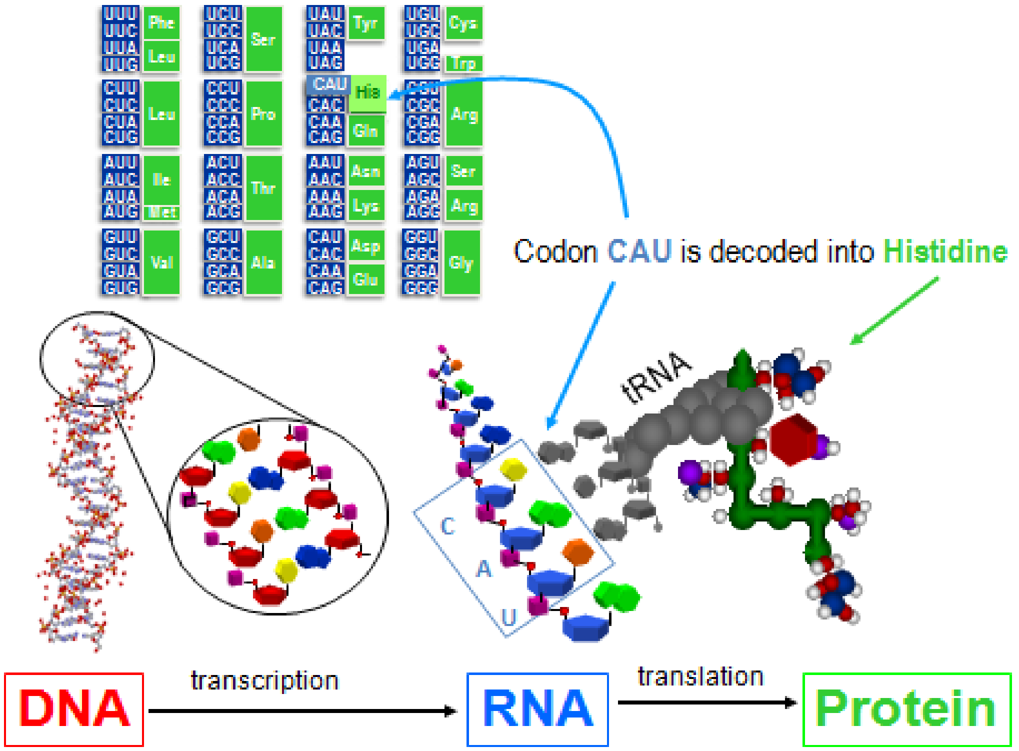


Figure 1. For at least 3 billion years, all living systems have consisted of a process by "translating" genetic information, "written" in the form of nucleotide sequences (chemically bonded into a linear chain known as nucleic acid) into corresponding sequences of amino acids (again chemically bonded together into a linear chain). A given sequence of amino acids will spontaneously fold up into a 3 dimensional shape, a protein that exhibits specific structural and functional properties, such as catalyzing a specific chemical reaction. Networks of protein molecules thus interact to form life – including the processes by which nucleic acid information is copied into a new generation through reproduction.

The genetic code nothing more and nothing less than the set of rules defining exactly what each possible genetic code-word (codon) is translated into. All codons are exactly 3 nucleotides in length, thus the genetic code is usually depicted as a table in which each possible codon (3 letters of RNA – a minor chemical derivative of the more famous DNA).

The genetic code is put into effect by a class of molecules known as tRNA's. Each tRNA is an adaptor molecule: at one end of the tRNA, an "anti-codon" recognizes a specific codon through chemical interactions; at the other end, the tRNA carries a specific amino acid. In this way, a suite of tRNA molecules enable translation of a specific gene sequence into a correspondingly specific sequence of amino acids.

Quite beyond this impact on human knowledge, the legacy of the central dogma awarded a far greater importance to the microbial lineage(s) that first evolved genetic coding, the organisms at the forefront of this historical transition. Current data indicate that the translation of genetic information into proteins, with all its language-like qualities, appeared more than 3 billion years ago (that is, around three quarters of our planet's history) (Fox, 2010). The result was a system of such evolutionary plasticity that biologists of the twenty-first century are struggling to find physical conditions that represent boundaries to life(Des Marais *et al.*, 2008). Already so-called extremophile organisms are known to have adapted to life at the deepest depths of the ocean and the highest reaches of the atmosphere (Satyanarayana, Raghukumar and Shivaji 2005). They divide and multiply in boiling water, deep-frozen ice, lakes of acid and streams of alkali. They survive extra-terrestrial trips on our space-craft and even grow on spent nuclear reactor fuel. It is not difficult to understand why luminaries of evolutionary biology list the advent of genetic coding as one of the major evolutionary transitions for life (Maynard Smith and Szathmary, 1995), ranked alongside life's origin and the emergence of language for its contribution to our current existence. Understandably then, research into the origins of genetic coding is a well-populated field with a correspondingly large scientific literature. Yet despite the linguistic terminology used by biologists, the language-like properties of genetic coding rarely merit serious consideration – or even an overt acknowledgment that something here deserves exploration. Is the trans-disciplinary vocabulary meaningful or misleading? Are we creating potential confusion with a careless analogy, or have we identified a useful metaphor that allows useful transport of insights and ideas across a large interdisciplinary divide?

Genetic coding and arbitrary signifiers

The key to any form of translation, including that found at the heart of biochemistry, is a code. A code is a system of rule(s) for converting one piece of information (for example, a letter, word, phrase, or gesture) into another representation (see for example Oxford English Dictionary, online edn). In other words, codes are interfaces that allow semantic information (useful meaning) to flow between languages. For example, when we talk about the Morse Code, we refer to the system of rules that describes how to transfer meaning between a simple alphabet of dots-and-dashes into the richer language of letters and symbols with which this paper is written. Likewise, the genetic code is the interface between a simple alphabet of 4 nucleotides and the richer alphabet of 20 amino acids (Figure 1). Just as Morse code assigns code-words constructed from the simple alphabet to a single letter of the more complex alphabet (for example *dot-dot-dot* is the code-word assigned to the letter *S*), genetic coding assigns nucleotide triplets, known as codons, to a single amino acid (such as *A-A-A* used to encode the amino acid *Lysine*).

In this sense, popularized science feeds confusion when it uses the phrase "genetic code" to refer to the genetic information stored within DNA (see, for example, "*Cracking the Genetic Code of Brown Tides*," Rudolf, 2011)). DNA is *not* a genetic code: rather, the genetic code is what imbues DNA with useful information. The information content of DNA (or any other language) is entirely defined by the code at work. Whether we consider a gene written in DNA, a gesture passed between a dog and its owner or your interpretation of this written sentence, changing the code changes the information. At the risk of labouring this crucial point, consider the symbols:

IOI

If we are decoding Arabic numerals into the concept of a number, this could represent "one hundred and one" (under a decimal notation) or "nine" (assuming binary notation). Removing the assumption that decoding should result in a number, text-processing software using the ASCII code would perceive a binary number signifying "horizontal tab". Dropping any assumption that numbers are involved, the symbols might represent a text message for "laugh out loud" or even the sigil of the comic book superhero Green Lantern. Semantic information is defined by the code.

From this standpoint, a fundamental question to ask of any code is the extent to which it is arbitrary. In other words, are the rules that connect two different representations of information constrained to reflect anything inherent to one or both of those representations, or to the process of translation itself? Drawing from linguistics, we suggest a metaphor of the connections between words and their meanings. At one extreme, words can be onomatopoeia: here, a non-arbitrary connection is implied as the word emulates its meaning directly. At the other extreme, words can be entirely arbitrary signifiers of the meanings they signify.

In contrast to the history of structuralism in linguistics (which seems to have *assumed* that arbitrary signifiers form the basis of language), early researchers of the genetic code *assumed* that a type of molecular onomatopoeia must be at work. Somehow, sequences of nucleotides evolved to encode amino acids. In the absence of intelligent design, how could arbitrary interactions arise? The clear alternative was to imagine direct interactions between the types of molecule involved – nucleotides and amino acids. For example, the famous double helix of DNA is a molecule with a distinct, physical structure that presents a surface rich with texture and associated chemical properties (such as positively and negatively charged atoms). Amino acids are likewise precise molecular structures with their own distinct shapes and chemical properties. An intuitive guess was that amino acids bound to DNA through direct, complementary chemical affinities (Gamow, 1954). Great scientific minds of the mid twentieth century devised ingenious models to describe such affinities in ways that would form genetic coding as a direct by-product of physics and chemistry (reviewed with clarity and wit by Hayes, 1998).

The problem with all such onomatopoeia-like theories of direct-templating was the subsequent discovery of a type of molecule known as transfer RNA (usually abbreviated to tRNA). All organisms contain a set of slightly different tRNA molecules that are the physical manifestation of the rules of genetic coding (i.e. the genetic code). At one end, each tRNA recognizes and binds to a specific codon (or, sometimes, a suite of similar codons); at the other end, it carries a specific amino acid. The very existence of this molecule, inserted between the genetic information and its amino acid counterpart, removes any need for direct physical complementarities between nucleotides and amino acids; relationships specified by the genetic code are as (potentially) different as the association between words and their meanings. Put another way, codons (and the nucleotides of which they are composed) are, in the modern system, symbolic signifiers for the amino acids they encode.

By the late 1950's, this much was known (Hoagland, Stephenson, Scott, Hecht & Zamecnik, 1958), and the concept of a truly symbolic code is important to much of what follows below. However, we would be remiss in finishing this section without reference to much more recent findings which have resurrected the concept of direct-templating (molecular onomatopoeia). During the late 1980's, biotechnologists invented a method known as SELEX (Tuerk & Gold, 1990). The process uses *in vitro* selection to find RNA molecules known as *aptamers* that target and bind other molecules (Cho, Lee & Ellington, 2009). When multiple aptamers are selected for their propensity to bind a specific target molecule, researchers often discover common motifs occur within them all. Since individual aptamers share little or no common ancestry, these common motifs are likely responsible for their binding affinities. Interestingly, statistical analyses reveal that aptamers selected for affinity to bind specific amino acids contain motifs that are unusually enriched with the codons assigned to those amino acids in the genetic code (Yarus, Widmann & Knight, 2009). Extending our previous metaphor, it would be as if cultural anthropologists found numerous tribes had independently derived vocabulary for a dog's bark and had all arrived on minor phonemic variations of "woof".

Biochemically, this intriguing result brings as many questions as answers. So far, only around half of the 20 amino acids have been tested with SELEX (and already one has shown no interesting associations). This incomplete data set renders any firm conclusions difficult. More generally, the lesson from early, ingenious theorizing about the code is that almost any pattern can been identified if one looks hard enough. One particularly poignant example concerns a 1966 publication which used plastic models of the relevant atomic structures to reveal an excellent fit between codons and their amino acids (Welton & Pelc, 1966). Critical re-evaluation revealed, however, that the researchers "had built all their polynucleotide sequences backwards. Their AAG was, in fact, GAA" (Crick, 1967) which encodes an entirely different amino acid of entirely different shape and properties from the amino acid for which an excellent fit had been discovered.

Most mysterious is why any putative direct-templating associations should have survived the evolutionary transition to tRNA-based, symbolic coding. The insertion of adaptor molecules would clearly disrupt any mechanism of direct interactions between amino acids and their encodings. Stretching our metaphor further, it is as if human communication began with nothing but onomatopoeic utterances – and yet in some distant future, long after all human communication evolved to indirect exchange between cell phones and computers, we found these primitive onomatopoeia operating unscathed. One interpretation is that changes to the rules of a code are impossibly disruptive once it has become established – no matter how much machinery evolves to separate signifiers from the signified. This idea seems logically attractive – but it also introduces our second point of exploration for similarities between biochemical and human language acquisition: what do we know about the emergence of a standard genetic code, shared by organisms as diverse as humans and *E. coli* bacterium?

Universal codes and footprints of evolution

A common challenge for understanding the evolution of any biological system is to find reliable clues that lead from the world we can observe backwards into the past. Broadly speaking, evolutionary research has developed two complementary approaches to find these footprints of evolutionary history, both of which present interesting parallels within the study of language.

The first approach seeks non-randomness in the form of apparent design in order to address the question, "what is this organism adapted for?" For example, we can deduce that insects which resemble twigs have been shaped by evolution to reduce their chances of being eaten by predators. More accurately, we infer that within the ancestral population from which these organisms derive, individuals with the greatest resemblance to twigs had a statistically higher chance of surviving to reproduce. Given time and the constant trickle of new, random mutations these genetic advantages accumulated into the startlingly non-random coloration and shape that first drew biologists' attention. Of course, this much alone is mere hypothesis, but it suggests a barrage of possible tests (do variations in natural insect populations correlate camouflage with survival and reproduction? Can we artificially modify insects' camouflage in order to measure the effects on predation?) In this manner, apparently purposeful design initiates evolutionary inquiries that end up revealing the evolutionary forces or 'selective pressures' from which a modern form has derived.

A less widely appreciated alternative approach searches contemporary structures for precisely the non-random features that do not make sense as adaptations – because these are often clues to an ancestral state. Specifically, features of an organism that do not contribute to an adaptation (or, better yet, those that compromise the perfection of an adaptation) can often be understood as restrictions imposed by the ancestral states from which an organism has evolved. Thus our own species' vulnerability to back and knee injuries reflect natural selection for bipedal locomotion that improvised on an anatomy previously adapted for quadrupedal gait.

Each approach has played an important role in shaping the present state of knowledge for the evolution of genetic coding. In the first category, the earliest suggestions for adaptive, non-random features of genetic coding focused on the fact that the code seems to have a built in mechanism for reducing the impact of errors. Codons that differ only at the third nucleotide position are often assigned to the same amino acid, for example codons GGU, GGC, GGA and GGC are all assigned to the amino acid *glycine* (Figure 1). Early researchers suggested this feature could have arisen from natural selection to minimize the impact of errors during translation because a high proportion of mistakes involving a single "letter" of the triplet codon have no effect on which amino acid is specified (Sonneborn 1965, Zuckerkandl & Pauling 1965). In other words, we might imagine a primordial population in which different genetic codes were at work. If occasional errors in translation were inevitable, then successful lineages might have been those operating genetic codes which mitigated the associated negative impact. That is, error minimising codes would be like well-camouflaged insects, using their adaptive advantage to out-survive and therefore out-reproduce their competitors.

This satisfyingly simple evolutionary argument was soon silenced by another discovery of Francis Crick. Specifically, his "wobble hypothesis" noted that tRNA adaptor molecules are often unable to distinguish between similar nucleotides at the third position of a codon for reasons of physics and chemistry (Crick, 1966). Under this view, the pattern of synonymous codons is a necessary byproduct of physical limitations rather than a clue to evolutionary adaptation. Crick's insight combined with his own, earlier flawed hypotheses (Crick, Griffith & Orgel 1957) to convince him that altogether too much excitable speculation had accompanied scientific discovery of the genetic code. His highly influential voice thus went on to assert that the code was best understood as a "frozen accident" of evolutionary history (Crick, 1968). To avoid mis-representing this claim, these are his exact words:

"This theory states that the code is universal because at the present time any change would e lethal. This is because in all organisms ... the code determines ... the amino acid sequences of so many highly evolved protein molecules that any change ... would be highly disadvantageous unless accompanied by many simultaneous mutations to correct these mistakes produced by altering the code. This accounts for the fact that the code does not change. To account for it being the same in all organisms, one must assume that all life evolved from a single organism (more accurately from a small interbreeding population). **In its extreme form, the theory implies that the allocation of codons to amino acids ... was entirely a matter of chance**" (page 370, emphasis added).

It is here that we re-connect to the evolution of human language. The Frozen Accident hypothesis and its effect in silencing evolutionary hypotheses about the origin of genetic coding seem to us oddly similar to the views expressed by another academic giant, Noam Chomsky, on universal grammar and their impact on research into the evolution of language acquisition (Chomsky, 1988). Both Crick and Chomsky concluded that because of the universality of the phenomenon (the genetic code and the structure of human language respectively), any attempt to infer evolutionary history is merely unhelpful speculation.

Genetic code research has escaped from such stifling reasoning due largely to the discovery of empirical, contradictory evidence. In fact, the first example of a naturally occurring variation in the genetic code was observed in none other than our own species. All cells of the human body contain mitochondria – organelles that house their own genetic material (and operate their own genetic codes) independent of the main cell nucleus. A decade after the frozen accident was put forwards, human mitochondria were found to use a code in which two codons are assigned to different amino acids than those described by the "universal" genetic code (Barrell, Bankier & Drouin, 1979). Since then, many other exceptions have been reported in a wide variety of species (Knight, Freeland & Landweber, 2001). This has eroded credibility of the frozen accident while reinvigorating the search for adaptive features that hint at the evolutionary forces by which genetic coding emerged. Particular success has come from resurrecting and expanding early arguments for an "error minimising" code. Indeed, sufficient evidence has accumulated for this concept that a paper of this length would be required to provide an adequate overview. Happily, such reviews already exist, both in specialized academic publications (Freeland, Wu & Keulman 2003; Koonin & Novozhilov 2009) and written for interested non-specialists (Freeland & Hurst, 2004). For present purposes it suffices to give a flavour of the argument in Figure 2.

The message is that even with the simplest quantification of non-randomness, the standard (non-universal) genetic code appears quite extraordinary compared to theoretically plausible alternatives. At first sight, finding a highly optimized pattern of codon arrangements might seem exactly the sort of evolutionary overlay that would disrupt the molecular onomatopoeia of a primordial, direct-templating system. One important study offers a resolution to the paradoxical evidence that both ideas are true by showing that the vast range of plausible genetic codes offers plenty of room for both ideas: many codes could fulfill the restrictions of onomatopoeia without displaying "error minimization" properties (Caporaso, Yarus & Knight, 2005). Even the simple "language" of genetic coding contains enough flexibility to carry more than one strong signal of its evolutionary causation.

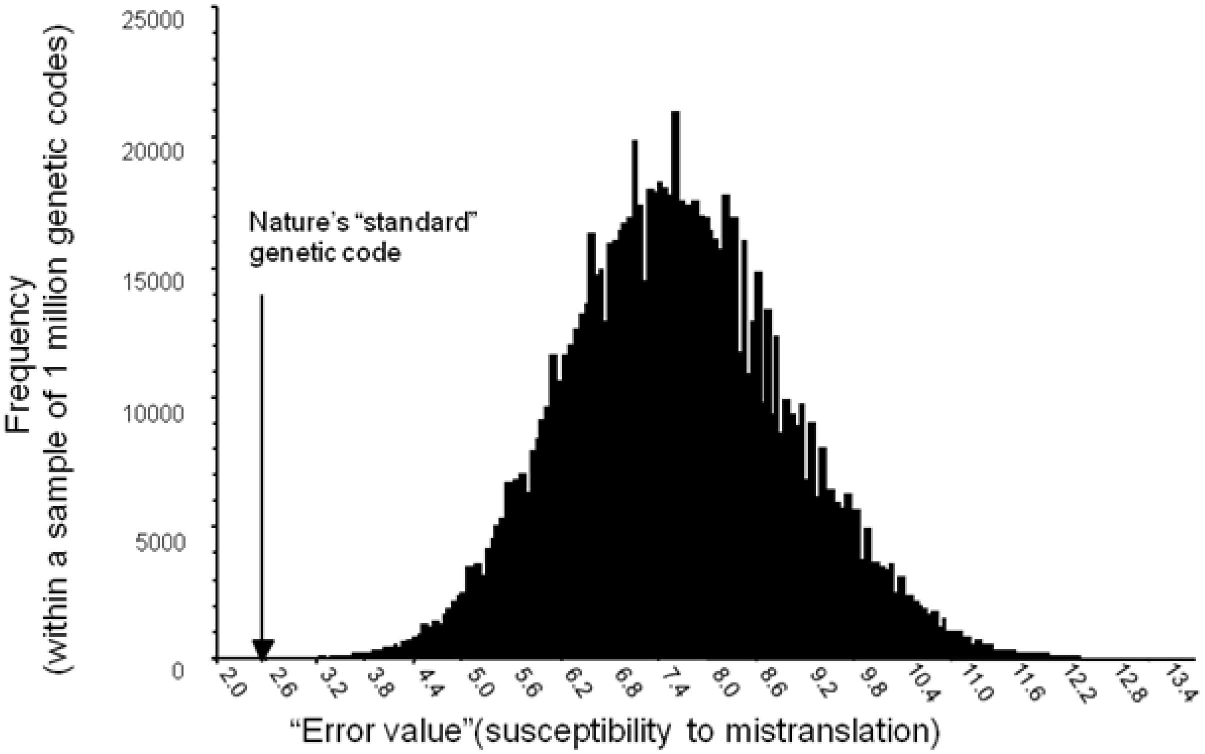


Figure 2. The standard genetic code is highly unusual by virtue of minimizing the impact of common errors that cause mistranslation. The standard genetic code is the particular set of rules that had evolved by the time of the last universal ancestor of life on Earth (and remained largely unchanged ever since).This figure illustrates a typical analysis of the standard genetic code compared to a computer-generated sample of 1 million randomized alternative codes, each considered plausible by maintaining all components of the molecular machinery used by the "real" code. Each genetic code is reduced to an "error value" that represents the code's susceptibility/ sensitivity to translation errors. The most common form of mistranslation event involves mis-reading a single nucleotide "letter" within the triplet that forms a codon. This can have a variety of effects on the fitness of an organism, according to the chemical similarity of the amino acid pair that forms the "actual" and "intended" translation of that genetic code word. A code with a large error values is one for which the average effect of common translation errors leads to a substantial change in the amino acids' chemical properties. A genetic code with a small error value is one which buffers the living system from this effect. In this particular study, within the sample of 1 million plausible alternative codes, only one was found that (slightly) improves upon the rules of the standard code in these terms.

Until recently, definitions of plausible alternative codes have focused on rearranging the pattern by which codons are assigned to 20 amino acids. This matches the empirical observations of non-standard codes that occur in nature (Knight *et al*. 2001), and increasingly robust insights into how codon reassignment takes place (Rocha, Pereira, Santos & Macedo-Ribeiro, 2011). More recently, however, a 21st and 22nd amino acid have been identified mid-way through incorporation into the genetic code, each by a different mechanism (Yuan, O'Donoghue, Ambrogelly, Gundllapalli, Sherrer, Palioura, Simonović & Söll, 2010). This in turn has contributed to deeper inquiry as to why evolution might "choose" a particular suite of 20 amino acids from a much larger pool of plausible alternatives. Once again, the startling finding is just how clearly biology choice distinguishes from any reasonable expectation of random chance (Freeland & Philip, 2011).

From our perspective, efforts to understand the evolution of genetic coding by identifying adaptations exhibit clear conceptual parallels to the foundational work of linguists who have identified similar explanations for the evolution of phonemic alphabets in vowel systems (Liljencrants & Lindblom, 1972; Lindblom, 1986). The primary difference seems to be one of empirical evidence: the recognition awarded to non-standard genetic codes is not yet matched by widespread acceptance of variations to "universal grammar." Cogent discussion elsewhere argues why this absence of evidence is entirely unsatisfactory both philosophically and scientifically as a justification for avoiding the question of where such a universal grammar has come from (MacNeilage, 2008: Chapter 1, pages 3-28), and we hope this is one area where the discoveries concerning the evolution of biochemical "language" will encourage those who persevere with evolutionary investigations to the acquisition of human language. Indeed, much of MacNeilage's (2008) approach to studying the mechanistic origins of human language parallels the other major approach to deriving clues to the history of genetic coding – the search for non-randomness that makes no sense in terms of adaptation. Here the undoubted pioneer of such thinking is Wong, who has worked for more than three decades to show that the standard genetic code can be traced back to earlier, simpler ancestral forms (which encoded fewer amino acids). A cornerstone of this argument is that only around half of the 20 amino acids used in modern genetic coding were available to early life (contemporary data for this point of view reviewed in Higgs & Pudritz, 2009 and Cleaves, 2010): the rest were invented by early evolution and then incorporated into the genetic code. As Wong points out, the sequence of biochemical steps or 'biosynthetic pathways' by which living organisms still synthesize each of the 20 amino acids seems to reflect this process: amino acids which appeared late within genetic code evolution are still made as chemical adjustments to amino acids that were likely available to the origin of life(Wong, 1975). Moreover, it appears that late-arriving amino acids may have entered the genetic code by capturing a sub-set of the codons previously assigned to their "early" counterparts, leaving behind patterns that make no sense in terms of error minimization. Once again, there is much more to this argument than we have space to discuss – and we again refer readers to one of several reviews that Wong offers to his own work (Wong, 2003). The important point is simply that in addition to its adaptive features and direct-templating origins, the code seems to carry a non-adaptive footprint of its evolutionary history. In this light, we concur strongly with the view (McNeilage, 2008: Chapter 2, especially pages 50-58) that it is strange, to say the very least, to claim that because language is a complex phenomenon, it self-evidently offers no quarter for meaningful evolutionary investigation (e.g. Chomsky, 1988 p. 170; Chomsky 2000, p. 4; p 49).

Origins of a generative grammar?

A current, widespread view from biology is that even the origin of genetic coding is logically distinct from the origin of genetic information. In particular, contemporary scientific thought favours the notion of an "RNA world," according to which the central dogma of molecular biology (i.e. the world of genetically encoded proteins) was preceded by one in which a single type of molecule, RNA, fulfilled all biological functions.

The RNA-world hypothesis derives support from two major sources. The first of these began with a Nobel prize-winning demonstration that RNA can spontaneously fold into shapes that catalyze specific biochemical reactions (1989, prize in chemistry). Subsequently, the same technology that has produced evidence for direct templating between RNA and amino acids (section "Genetic coding and arbitrary signifiers") allowed other researchers to uncover an unsuspected wealth of protein-like activities for other RNA molecules (Atkins, Gesteland and Cech, 2011). They too can fold, according to their sequence, into complex shapes that catalyze specific chemical reactions – indeed RNA molecules that do are known as ribozymes to mark their functional equivalence to protein enzymes.

A second, complementary source of support for the RNA-world hypothesis can be traced back to those who first deciphered genetic coding, including (who else?) Francis Crick. Their insightful observation was that the key molecular components of the genetic code were in fact not proteins at all. For example, in the words of Crick(1966, page 7), "*It almost appears as if tRNA were Nature's attempt to make an RNA molecule play the role of a protein*" and "*It is tempting to wonder if … the primitive machinery had* *no protein at all and consisted entirely of RNA*“ (Crick, 1968, page 371) These observations combined with identification of other, non-informational roles played by nucleotides in modern metabolism to create a concept “molecular fossils” (White 1976): molecules found in living systems unchanged from their ancient evolutionary debut. Technological breakthroughs of the gene-sequencing culture have allowed scientists to study specific molecules with enough detail to build an evolutionary paradigm expressed by another language metaphor: that modern metabolism is best understood as an evolutionary palimpsest - a document on which one layer of writing has been largely erased to make way for a second, later inscription (Benner, Ellington and Tauer, 1986). According to this view, a primordial metabolism constructed with interacting RNA molecules has been largely ‘over-written’ by a secondary layer of protein molecules since the advent of genetic coding.

We resist this current, widespread view and instead assert that neither line of evidence can support the RNA-world hypothesis over its exact conceptual opposite: a primordial, self-replicating metabolism of proteins which later “invented” RNA as genetic material. Our (distinctly minority) view is most easily explained in terms of the palimpsest metaphor. For a real palimpsest, the only way in which a chronology of two written layers is with reference to a third layer: the parchment on which the writing appears. Since biology contains no analogue for this writing-surface, nothing can tell us which “layer” came first. In other words, the existence of RNA molecules performing protein-like roles could equally well indicate that RNA evolved to usurp functions previously given over to proteins. RNA would have been capable of escaping from the role of genetic information storage precisely because of its incidental capacity to fold, form shapes and catalyze chemical reactions.

We introduce this unusual argument because this is where we see the most immediate potential for linguistic theory to inform ideas of biochemical evolution. Within linguistic theory, we are led to understand that anyone who does ask questions about the origins of generative grammar perceives that this generative grammar clearly emerged as an evolutionary refinement of some earlier, heterogeneous mixture of sounds and gestures that lacked a coherent, standardized framework. We further understand that this view is supported by a well-established body of appropriate logic (Pagin). We doubt that many evolutionary biologists would fail to see the logical difficulties (teleology) of supposing that a regularized information-storage system evolved in order to be ready for the subsequent evolution of content when presented with that clarity of thinking. However, most evolutionary biology was derived to describe the “modern” world in which the generative grammar of nucleic acid genes already exists. Here, nucleotide sequences are indeed the legitimate focal point for understanding most evolutionary outcomes, precisely because they are the physical basis of heredity (e.g. see Dawkins 1976). The careless extrapolation is to assume that nucleotide sequences have always played this role – that the origin of genetic information is one and the same thing as the origin of nucleotide sequences. From beyond the traditional disciplinary boundaries of evolutionary biology, this assumption seems questionable. For example, chemists widely agree that even individual nucleotides (the building blocks of RNA), let alone sequences of them joined together, are harder to form under pre-biological conditions than almost any other biologically relevant molecule; they also agree that amino acids are amongst the easiest. Astronomy agrees, having found no trace of nucleotides beyond Earth, but plenty of evidence to suggest that amino acids (and other simple organic compounds) quite literal rain down on the surface of every planet in the cosmos (Sephton). Meanwhile, biochemistry has long known that aside from their modern role as building blocks for genes, nucleotides play many other non-informational roles that happen to show every conceivable hallmark of “molecular fossil” status (such storing energy that is subsequently released to power protein-based metabolism). These are at best mere hints for an alternative view of the origins of genetic information – but they are consistent with something deep that linguistic theory seems to have already learned: that generative grammar is a refinement of an earlier, less modular system. If our previous discussion of genetic code evolution is intended to encourage those who study the evolution of human language, then this final comment reverses the relationship: it seems relevant and important to ask our evolutionary linguist colleagues to guide us through their understanding of the emergence of generative grammar to see what clarification it can bring to the study of life’s origins.

Concluding remarks

In this paper we provide a brief overview of three themes relating to the broad topic evolution of biochemical "language": the origin of genetic coding, the emergence of the standard (not universal!) genetic code, and the origin of genetic information. The sequence in which we introduce these topics is deliberately chosen to present a progression from vague (and perhaps careless) use of linguistic analogy into increasingly direct suggestions of meaningful similarities. We are fully cognizant that metaphors are not to be mistaken for literal truths simply because the results would be interesting – but we consider ongoing exploration to be important. At the very least, if all linguistic constructions are metaphors for the concepts they represent, then all scientific explanations carry with them the strengths and weaknesses inherent to metaphor. They can illuminate and clarify our thinking or mislead us with the incidental baggage they bring along. Greater awareness of this point alone cannot harm scientists in their thinking, but we believe that far more is at stake.

In our introduction we identified ourselves as astrobiologists: scientists who actively seek to place their research and thinking within a broader framework of thought that defines life's place within the cosmos. The inherent interdisciplinarity of astrobiology has already yielded profound and unexpected insights. Placing the findings of astronomy next to those of chemistry and evolutionary biology reveals that half of the amino acids at work in your genetic code were likely brought to Earth by comets (reviewed in Cleaves, 2010; Higgs and Pudritz 2009). Placing the findings of geochemistry and planetary science next to those of biology reveals that life has changed our planet every bit as much as our planet has defined the environment for life (for example, the existence of an oxygen-rich atmosphere is a product of early life, not a pre-requisite for life (Kasting 2001); as much as half of Earth's minerals owe their very existence to biology (Hazen 2010). Overarching such individual insights, an overwhelming majority of scientific findings within the past fifty years suggest that life should be more common within our galaxy than was previously thought. For example, the existence of an extra-solar planet (i.e. a solar system other than our own) was confirmed as recently as 1995; the number of exoplanets known to science has been jumping ever since (for a recent estimate, see Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., Wambsganss, J., Menzies, J. *et al* 2012); life on our planet seems to have started much earlier than was previously thought (Mojzsis, 2002), and life appears to tolerate a much larger range of conditions than anyone guessed (Satyanarayana et al. 2005). All of this places an increasing significance on the importance of understanding whether life is destined by physics to converge, through evolution, towards predictable outcomes. At present, it remains ambiguous whether and where the superficial similarities of human language and biochemical "language" indicate something fundamental about the universe – about the emergence of generative grammars (combinatorial representation systems), about the way in which symbolic coding can emerge, and what signs of their evolution they leave behind. This would do much to inform our expectations about what characteristics we might legitimately expect from an independent origin of life. More parochially, it offers the potential to understand more about the relative roles of chance versus inevitability in steps that are fundamental to our own existence.

Acknowledgements

We are extremely grateful to Bjorn Lindblom and Francisco Lacerda for hosting the conference that led to this manuscript, to Peter MacNeilage and Peter Pagin for patience in explaining to us the fundamental ideas of evolutionary thinking within linguistic theory. We also thank Karen Meech for granting us time to pursue this particular interdisciplinary conversation. Sara Walker, Andrew Pohorille and James Stephenson have encouraged us in an exploration of alternatives to the RNA world hypothesis, and the NASA Astrobiology Institute has offered persistent encouragement for all interdisciplinary interactions that bring new insights to the study of life's origins. This material is based upon work supported by the National Aeronautics and Space Administration through the NASA Astrobiology Institute under Cooperative Agreement No. NNA09DA77A issued through the Office of Space Science.

References

Armstrong, K. (1993). A History of God. (*Ballantine Books*). Chapter 6 “The God of the Philosophers”, pp. 170-208.

Barrell, B. G., Bankier, A. T. & Drouin, J. (1979). A different genetic code in human mitochondria *Nature* 282: pp. 189 – 194.

Benner, S., Ellington, A.D. & Tauer, A. (1989). Modern metabolism as a palimpsest of the RNA world. *Proc. Nati. Acad. Sci. USA* 86: pp. 7054-7058.

Caporaso, J.G., Yarus, M. & Knight, R. (2005). Error Minimization and Coding Triplet/Binding Site Associations Are Independent Features of the Canonical Genetic Code. *J Mol Evol* 61:597–607.

Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., Wambsganss, J., Menzies, J. et al (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature 481*: pp. 167–169.

Cho, E.J., Lee, J.W. & Ellington, A.D. (2009). Applications of Aptamers as Sensor. *Annual Review of Analytical Chemistry* 2: 241–64.

Chomsky, N. (1988). *Language and the problems of knowledge.* (Oxford: Oxford University Press.).

Chomsky, N. (2000). *The architecture of language*. (Oxford: Oxford University Press.).

Cleaves, H.J., II. (2010). The origin of the biologically coded amino acids. *J Theor Biol* 263:490–498.

Crick, F. (1966). Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 19: 548–55.

Crick, F. (1966). The Genetic Code –Yesterday, Today, and Tomorrow. *Cold Spring Harbor Symposia on Quantitative Biology* 31: 3-9.

Crick, F. H. C. (1967). An Error in Model Building. *Nature* 213: page 798.

Crick, F.H., Griffith, J.S., Orgel, LE. (1957). Codes without commas. *Proc Natl Acad Sci U S A* 43:416-21.

Crick, F.H.C. (1958). On Protein Synthesis. *Symp. Soc. Exp. Biol.* 12: 139-163.

Crick, F.H.C. (1968) .The Origin of the Genetic Code, *J. Mol. Biol.* 38: 367-379.

Des Marais, D., Nuth, J.A., Allamandola, L.J., Boss, A.P., Farmer, J.D., Hoehler, T.M., Jakosky, B.M., Meadows, V.S., Pohorille, A., Runnegar, B. & Spormann, A.M. (2008). The NASA Astrobiology Roadmap, *Astrobiology*, 8(4): pp. 715-730.

Filkins, P. (2012). How translation shapes meaning. *The Chronicle of Higher Education*, 58(28) pp. B12-B16.

Fox, G.E. (2010). Origin and evolution of the ribosome. *Cold Spring Harbor Perspectives in Biology* 2: pp. 1 – 18.

Freeland, S. & Philip G.K., (2010). Did evolution select a non-random “alphabet” of amino acids? *Astrobiology* 11: pp. 235–240.

Freeland, S. (2007). Could an intelligent alien predict earth’s biochemistry? In *Fitness of the Cosmos for Life* (eds. J. Barrow, S. Conway-Morris and S.J. Freeland), Cambridge University Press, Cambridge.

Freeland, S.J. & Hurst, L. (2004). Evolution Encoded, *Scientific American* 290:84-91.

Freeland, S.J., Wu, T. & Keulmann, N. (2003). The case for an Error Minimizing Standard Genetic Code., *Origins of Life and Evolution of the Biosphere* 33: 457-477.

Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. *Nature*, 173: p. 318.

Gould, S.J.(1989). *Wonderful Life: the Burgess Shale and the Nature of History*. (W.H. Norton and Company, New York and London).

Hayes, B. (1998). The invention of the genetic code. *American Scientist*, 86: pp. 8-14.

Hazen, R.M. (2010). Evolution of minerals. *Scientific American* 303: pp. 58-65.

Higgs, P.G. & Pudritz, R.E. (2009). A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* 9:483–490.

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. (1958). A Soluble Ribonucleic Acid Intermediate in Protein Synthesis *J. Biol. Chem.* 231: pp. 241–257.

Hughes, S. (2001). Making dollars out of DNA. The first major patent in biotechnology and the commercialization of molecular biology, 1974-1980. *Isis; an international review devoted to the history of science and its cultural influences* 92: 541–575.

James F. Kasting (2001) The Rise of Atmospheric Oxygen. *Science* 293: pp. 819-820.

Knight, R.D., Freeland, S. J. & Landweber, L. F. (2001). "Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, 2:49-58.

Koonin, E.V. & Novozhilov, A.S. (2009). Origin and evolution of the genetic code: the universal enigma. *IUBMB Life*. 61:99-111.

Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: pp 839-862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In *Experimental phonology* (eds. J.J. Ohala amd J.J. Jaeger), pp. 13-14. Orlando, FL: Academic Press.

MacNeilage, P.F. (2008). *The origin of speech*. Oxford, Oxford University Press: pp 1-28.

Maxam, A.M. & Gilbert W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74(2).

Maynard Smith, J. & Szathmáry E. (1995). *The Major Transitions in Evolution.* (Oxford, England: Oxford University Press).

Mojzsis, S.J. (2002) Origin of Life: The First Fossils. In *Encyclopedia of Evolution*, edited by M. Pagel: Oxford University Press, Oxford, pp. 843-845.

Morris, S. C. (2003). *Life's Solution: Inevitable humans in a lonely universe* (Cambridge University Press, Cambridge UK).

Munger, S. (1999). Bible, Babel and Babble: The Foundations of Bible Translation (*International Bible Society, Colorado Springs*).

Oxford English Dictionary, online edn. 2E: http://oxforddictionaries.com/definition/code.

Rahman, M., "History of Genetic Engineering" by: http://www.ehow.com/video_5112340_history-genetic-engineering.html.

RNA Worlds: From Life's Origins to Diversity in Gene Regulation (eds. Atkins J. F., Gesteland F. And Cech T.R. Cold Spring Harbor Laboratory Press, America).

Rocha, R., Pereira, P.J., Santos, M.A. & Macedo-Ribeiro, S. (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. *Proc Natl Acad Sci U S A* 108:14091-6.

Rudolf, J. C. (2011). Cracking the Genetic Code of ‘Brown Tides, *New York Times*, February 22.

Satyanarayana, T., Raghukumar, C. & Shivaji, S. (2005). Extremophilic microbes: Diversity and perspectives. *Current Science* 89: 78–90.

Saussure, F. (1983). Course in General Linguistics. (Eds. Bally C. and Sechehaye A.: Trans. Roy Harris. La Salle, Illinois: Open Court).

Sonneborn, T.M. (1965). Degeneracy of the genetic code: extent, nature and genetic implications. (Academic Press, New York).

The Nobel Prize in Chemistry 1972. Nobelprize.org. 13 May 2012 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1972.

The Nobel Prize in Chemistry 1989. Nobelprize.org. 13 May 2012 http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1989/.

The Nobel Prize in Physiology or Medicine 1962. Nobelprize.org. 13 May 2012 http://www.nobelprize.org/nobel_prizes/medicine/laureates/1962/.

The Nobel Prize in Physiology or Medicine 1968. Nobelprize.org. 13 May 2012 http://www.nobelprize.org/nobel_prizes/medicine/laureates/1968/.

The Nobel Prize in Physiology or Medicine 1969. Nobelprize.org. 13 May http://www.nobelprize.org/nobel_prizes/medicine/laureates/1969/.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–51.

Welton, M.G.E. & Pelc, S.R. (1966). XXX *Nature* 209: page 868.

White, H.B. (1976). Coenzymes as Fossils of an Earlier Metabolic State *J.Mol.Evol.*7,101-104.

Williams, D., Fournier, G.P., Lapierre, P., Swithers, K.S., Green, A.G., Andam, C.P. & Gogarten, J.P. (2011). A rooted net of life. *Biol Direct.*

Wong JT. (2005). Coevolution theory of the genetic code at age thirty. *Bioessays*. 27:pp 416-25.

Wong, J. T.-F. (1975). A Co-evolution Theory of the Genetic Code, *Proc. Natl. Acad. Sci. USA* 72: pp 1909−1912.

Yarus, M., Widmann, J.J. & Knight, R. (2009). RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol.* 69:406-29.

Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Sherrer, R.L., Palioura, S., Simonović, M. & Söll D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett.* 584: pp. 342-9.

Zuckerkandl, E. & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. (*Evolving Genes and Proteins* Bryson V. & Vogel H. J. (eds), Academic Press, New York and London).

Self-Organization:
Complex Dynamical Systems in the Evolution of Speech

Pierre-Yves Oudeyer
INRIA Bordeaux Sud-Ouest
pierre-yves.oudeyer@inria.fr

Abstract

Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities – universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What is the origin of the forms of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions. In particular, a computational model is presented and shows that basic neural equipment for adaptive holistic vocal imitation, coupling directly motor and perceptual representations in the brain, can generate spontaneously shared combinatorial systems of vocalizations in a society of babbling individuals. Furthermore, we show how morphological and physiological innate constraints can interact with these self-organized mechanisms to account for both the formation of statistical regularities and diversity in vocalization systems.

Keywords: self-organization, natural selection, evolution of speech, combinatoriality, computational model.

Introduction

Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities – universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What are the origins of the forms of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions.

The tendency of many complex physical systems to generate spontaneously new and organized forms, such as ice crystals or galactic spirals, is indeed present as much in the inorganic world as in the living world. Thus, the explanation of the origins of forms and structures in the living can not only rely on the principle of natural selection, which should be complemented by the understanding of physical mechanisms of form generation in which self-organization plays a central role. This applies to the social and cultural forms of the living, in particular to the forms of speech and language. As a consequence, I will begin by articulating in a general manner the relationships between self-organization, natural selection and neo-Darwinism in explanations of the genesis of forms in the living. Then, I will instantiate these relations in the context of the three questions formulated above. After that, I will explain why the use of computer models and simulations is fundamental for the progress in our understanding of these issues. Finally, I will present the example of an experiment based on a computer model which shows that certain simple mechanisms coupling perception and production of sounds can generate combinatorial systems of vocalizations, characterized by the universal/diversity duality, and shared culturally by the members of a speech community. I will conclude by the presentation of evolutionary scenarios that this computer experiment complements or renews.

Self-Organization and the Evolution of Forms in the Living

Physics, the Caldron of Self-Organized Forms

Nature, especially inorganic nature, is full of fascinatingly organized forms and patterns. The silhouette of mountains is the same, whether one views it at the scale of a rock, a summit, or a whole mountain range. Sand dunes often arrange themselves in long parallel stripes. Water crystallizes into symmetrical serrated flakes when the temperature is right. And when water flows in rivers and hurtles over cataracts, trumpet-shaped vortices appear and the bubbles collect together in structures which are sometimes polyhedral. Lightning flashes draw plant-like branches in the sky. Alternating freezing and thawing of the rocky ground of the tundra leaves polygonal impressions in the earth. The list of these forms rivals many human artefacts in complexity, as can be seen in figure 1. And yet they are not designed or conceived by anyone or anything, not even natural selection, Dawkins' `blind watchmaker' (Dawkins, 1986). What, then, are the mysterious factors that explain their existence?

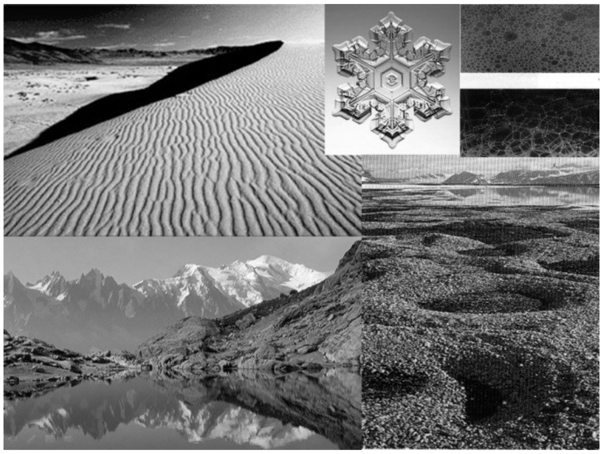
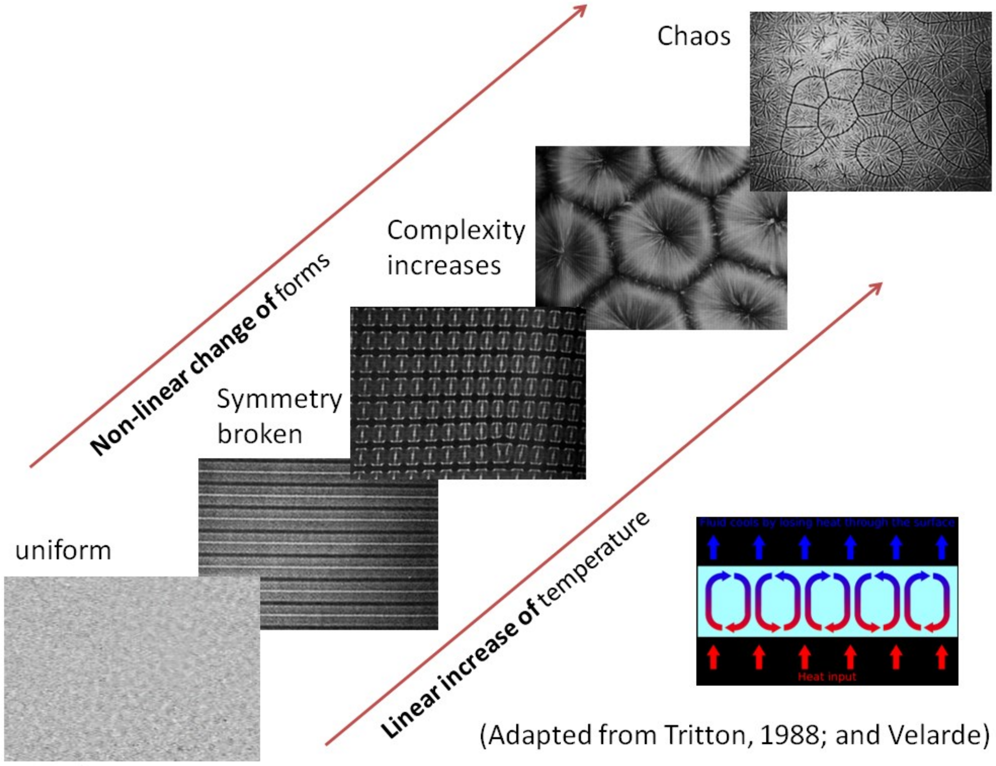


Figure 6. Nature is full of organized forms and patterns without there being anywhere any plans which might have served to build them; they are said to be self-organized. Here, parallel stripes running through sand dunes, water bubbles on the surface of liquid which has been stirred up and the polyhedral structures which are left when they dry out, an ice crystal, mountains whose shapes are the same whether one views them on the scale of a rock or a whole peak (Photos: Nick Lancaster, Desert Research Institute, Nevada).

In fact, all these organized structures have a feature in common: they are the macroscopic outcomes of local interactions between the many components of the system from which they emerge. Their global organizational properties are not to be found at the local level. Indeed the properties of the shape of a water molecule, as well as of its individual physico-chemical properties, are qualitatively different from the properties of ice crystals, whirlpools, or polyhedral bubbles. The polygonal impressions in the tundra do not correspond with the shape of the stones composing them, and have a spatial organization quite different from the temporal organization of freezing and thawing. This is the hallmark of a newly discovered phenomenon -- self-organization.

In nature, self-organization characterizes very diverse physical systems, but several typical properties can be identified: non-linearity, symmetry breaking, presence of dynamical systems “attractors”, and historicity. For example, when one heats from below a thin layer of oil spread out on a plane surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change dramatically when the temperature goes over given thresholds (see figure 2). On the contrary, between these thresholds the shapes remain globally stable even if they are perturbated, constituting attractors. Another property of many self-organized dynamical systems is historicity, often associated with the sensitivity to initial conditions in chaotic systems: the attractor in which the systems falls, i.e. the shapes/forms that are produced by the complex system, can be very different depending on slight variations in the initial conditions. For example, this is the case of ferromagnetization: each of the atoms of an iron plate can be viewed as a sort of magnet that can have several possible orientations, and which orientation is random if the temperature is high enough. Yet, is the temperature goes below a certain threshold, a self-organized phenomenon happens: all atoms spontaneously adopt the same magnetic orientation. This shared orientation is quasi-impossible to predict beforehand and tiny random variations of initial orientations of atoms can lead the plate to be magnetized in a quite different orientation. These variations in initial conditions are typically linked to contingent events that interacted with the iron plate: this is why the final state of the plate depends both on its history and on its intrinsic physical mechanisms, whence the term “historicity”.



(Adapted from Tritton, 1988; and Velarde)

Figure 7. Rayleigh-Bénard cells: when one heats from below a thin layer of oil spread out on a plane surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change brutally when the temperature goes over given thresholds. This kind of non-linearity characterized many self-organized systems both in the inorganic and in the living world24.

---

24 Photos adapted from (Tritton, 1988), and Manuel Velarde, Universidad Computense, Madrid.

This fundamental concept is the touchstone of the paradigm shift undergone by the sciences of complexity in the 20th century (Ashby, 1956; Nicolis and Prigogine, 1977; Kauffman, 1996; Ball, 2001). Ever since Newton, good science was supposed to be reductionist, and consisted in decomposing natural systems into simpler subsystems. For example, to understand the functioning of the human body, it was appropriate to study the respective parts, such as the heart, the nervous system, or the limbic system. Moreover, it did not stop there, and study of the nervous system, for example, was subdivided into study of the cortex, of the thalamus, or of the peripheral motor innervations, and each of these sub-parts was studied by hyper-specialists in separated dedicated university department. This method has obviously enabled us to accumulate an impressive bank of knowledge. But the prophets of complexity have broken up this paradigm. Their credo is ``the sum of the parts is greater than the parts taken independently''.

The Impact of Self-Organization on the Origins of Forms in the Living

Complex systems, i.e. systems composed of many interacting sub-systems, abound in nature and have the strong tendency to self-organize. The examples of the previous section were chosen deliberately from inorganic systems to show that the property of self-organization can be found in systems subject to laws which have nothing to do with natural selection. However, self-organization applies similarly to living systems. It is a concept widely used in several branches of biology. It is particularly central to theories which explain the capacity of insect societies to build nests or hives, to hunt in groups or to explore in a decentralized and effective way the food resources of their environment (Camazine et al., 2002). In developmental biology, it is used, for example, to explain the formation of coloured patterns on the skins of animals like butterflies, zebras, jaguars or ladybirds (Ball, 2001).

Thus, it seems possible, then, that there are shape- and pattern-forming mechanisms in biological systems which are orthogonal to natural selection, through their property of self-organization. Now natural selection is at the heart of almost all the arguments of biologists when it comes to explaining the presence of a shape, a pattern or a structure in an organism. What, then, is the relationship between the theory of natural selection and self-organization?

Some researchers have suggested that self-organization casts doubt on the centrality of natural selection in explaining the evolution of living organisms. Waldrop explains:

> *``Complex dynamical systems can sometimes go spontaneously from randomness to order; is this a driving force in evolution? Have we missed something about evolution - some key principle that has shaped the development of life in ways quite different from natural selection, genetic drift, and all the other mechanisms biologists have evoked over the years? ... Yes! And the missing element... is spontaneous self-organization: the tendency of complex dynamical systems to fall into an ordered state without any selection pressure whatsoever.'' (Waldrop, 1990)*

However, this is not the position I take in this article. Rather than seeing self-organization as a concept which minimizes the role of natural selection by suggesting competing form-creating mechanisms, it is more accurate to see it on the one hand as belonging to a somewhat different level of explanation and above all on the other hand as describing mechanisms which actually increase the power of natural selection by orders of magnitude. Mechanisms with the self-organizing property are completely compatible with the mechanism of natural selection in explaining the evolution of forms and structures in biology.

Classic Neo-Darwinism

To see the matter clearly, it is first necessary to recall what the mechanism of natural selection, or Neo-Darwinism, comprises. It is a mechanism characterizing a system composed of individuals each having particular traits, shapes or structures. In addition, the individuals in this system are capable of replication. This replication must occasionally produce individuals which are not exact copies of their ancestors, but are slight variants. These variations are the source of diversity among individuals. Finally, each individual has a greater or lesser capacity for replication, according to its surrounding environment. This generates differential replication of individuals and gives rise to ``selection'' of those who are most capable of replicating themselves. The combination of the processes of replication with variation and selection means that, over the generations, the structures or traits of individuals which help them to reproduce themselves are preserved and improved upon.

Now there is one crucial point on which the theory of natural selection is neutral: it is the way in which variation is generated, and more generally the ways in which the individuals with their shapes, traits and structures are produced. A number of Neo-Darwinist arguments consider the mechanisms of variation of forms as secondary in comparison with the reproductive advantages of these forms when it comes to explain their evolution. This implies implicitly that the relation between the levels of genes, considered as the main space in which variations operate through mutations and cross-overs, and the level of phenotypes, considered as an isomorphic image of the space of genes, is simple and linear. According to this vision, exploration of the space of phenotypes (which determines, along with the environment, the relative effectiveness of the genes at replicating) can simply be carried out by studying the way things change in the space of genotypes. Now the mechanisms of mutation which actually bring about these changes are of little amplitude (most mutations only affect a very small part of the genome when replication succeeds), and thus random variations in genes lead to uniform exploration of the space of genotypes. What this means is that under the hypothesis that phenotypic and genotypic space have the same structure and can be mapped approximately linearly, the space of possible biological forms can be searched quasi-continuously, by successive little modifications of pre-existing forms. Fortunately for the appearance of complex life-forms, this is not the case. In fact, although this mechanism of small successive variations in form is notably effective in the delicate regulation of the structures of organisms, it would make the search for forms as complex as those of human organisms equivalent to the search for a needle in a haystack because genomes are much too high-dimensional (Keefe and Szostak, 2001).

Self-Organization Constrains the Space of Forms to Be Explored: Not All Forms can Emerge Equally Easily.

It is here that the concept of self-organization comes to the rescue of this naive search mechanism in the space of phenotypic forms in the Neo-Darwinian theoretical framework. In fact the relation between genes and the forms of organisms is characterized by its complexity and its non-linearity that are expressed through the ontogenetic and epigenetic development of each organism. Organisms are constructed starting from a stem cell containing a whole genome. This stem cell can be seen as a dynamic system parameterized by its genome and under the influence of perturbations imposed by the environment. This dynamic system is crucially a self-organizing system with the same sorts of properties as the self-organizing systems described in the previous section. The genome is a set of parameters analogous to temperature and the viscosity of liquids in Bénard systems, and the environment is analogous to noise (but evidently highly structured noise!). Thus the development of an organism from a stem cell shares many properties with physical systems: shapes, structures and patterns appear at the global level, and are qualitatively different from those implementing functioning at the local level, that is, different from the patterns characterizing the structure of the stem cell and its genome. The hexagonal pattern which can appear as a result of a simple difference in temperature in a homogeneous liquid gives an idea of the way in which a simple sequence of nucleotides enclosed in a system of molecules which transforms them automatically into proteins can generate a bipedal organism endowed with two eyes and ears and an immensely complex brain.

Crucially, as with Bénard systems or ferromagnetic plates, dynamic systems defined by the cells and their genomes are characterized by a landscape of attractors: there are large regions in the parameter space within which the dynamic system systematically adopts behaviour which is more or less the same. For Bénard systems, there is a range of temperatures giving rise to parallel stripes which is wide enough to locate easily. For ferromagnetic plates the range of temperature in which the system settles to global magnetic coherence is also very wide. Thus for living organisms it is not only possible to generate self-organizing structures with complex global properties, but in addition these structures are generated by genomes belonging to broad sub-spaces of genome space, called basins of attraction. The structuring of genome space into basins of attraction by this kind of dynamic system facilitates the evolutionary search of the space of forms so that it does not resemble a search for a needle in a haystack.

As in ferromagnetic systems, structured noise imposed by the environment on the development of the dynamic system can lead it to take different developmental pathways. For pieces of iron at low temperatures, this corresponds to magnetization in one direction or another. For a living organism, this corresponds to its possible shapes; this is how it happens that even monozygotic twins can show quite important morphological differences. This is also the reason why the relationship between genes and the forms of organisms is not only complex and nonlinear, but also non-deterministic. Moreover, and as in Bénard systems where search of the parameter space of temperature can sometimes lead to fast and qualitative changes in the behaviour of the system (for example the change from parallel stripes to square cells), which have been called phase-transitions, the search within genome space can also lead to fast qualitative changes. This possibly corresponds to many observations of rapid form-changes in evolution, as witnessed by the fossils studied by anthropologists, and which are the basis of the theory of punctuated equilibrium proposed in (Eldredge and Gould, 1972).

To summarize, the self-organizing properties of the dynamic system composed by the cells and their DNA brings crucial structuring to the phenotypic space by constraining it, making the discovery of complex robust forms by natural selection much easier. On the one hand, these properties enable a genome to generate complex, highly organized forms without the need for precise specification of each detail in the genome (in the same way as Bénard's polygonal shapes are not specified precisely, or encoded in a plan, in the properties of the liquid's molecules). On the other hand, the self-organizing properties structure the landscape of these possible forms into basins of attraction within which they resemble each other greatly (here is where gradual evolution happens, involving fine tuning of existing structures), and between which there can be substantial differences among forms (transitions from one basin to another are what provide abrupt and powerful innovations in evolution). To give a simple picture, self-organization provides a catalogue of complex forms distributed over a landscape of valleys in which and between which natural selection moves and makes its choices: self-organization proposes, and natural selection disposes. Obviously this is only an image to facilitate understanding, because with its movements natural selection actually enables new mechanisms, themselves self-organized, to appear, and these in turn structure the space of forms within which it moves; thus natural selection participates in the formation of these mechanisms which help it to move effectively in the space of forms; vice versa, the mechanism of natural selection certainly appeared in the history of life due to the self-organized behaviour of systems which were as yet completely unconnected to natural selection; natural selection and self-organizing mechanisms thus help each other reciprocally in a sort of spiral which enables complexity to increase during the course of evolution.

The consequence of this entanglement between natural selection and self-organization is that any explanation of the origins and evolution of forms and structures in living organisms requires at least two kinds of argumentation. The first one, classic, is the neo-Darwinian functionalist argumentation: it consists in identifying the ecological context in which a given trait may have appeared and in articulating the balance between the associated costs and reproductive advantages. The second kind of argumentation is more rarely used but is equally essential: it consists in identifying the developmental/epigenetic mechanisms, and the associated constraints, which may have permitted, or made difficult, the genesis of these novel traits. And the concept of self-organization is central to the way developmental mechanisms impact the genesis of forms.

Self-Organization and the Evolution of Forms and Structures of Language and Languages

The question of how speech and language came to humans, and the question of how new languages form and evolve, are among the most difficult that science has to address. After being put aside from scientific enquiries during most of the 20th century, partly because of the ban pronounced by the Société Linguistique de Paris, they are now again the focus of attention of a whole scientific community. There is an emerging consensus among researchers who are today getting down to questions of the origin of the human language faculty and the evolution of languages -- this research must be interdisciplinary. It in fact poses a puzzle with immense ramifications which go beyond the competence of each individual discipline on its own. Firstly, it is because the two big questions, that of the origins of language and that of the origins of languages, must be decomposed into subquestions which are themselves already quite complex: What, in fact, is the language faculty? What is a language? How are sounds, words, sentences and representations of meaning related to each other? How does the brain represent and process these sounds and sentences and the concepts which they convey? How do we learn to speak? What are the respective roles of nature and nurture? What is language for? What is its role in a community? How does a language form and change in the course of successive generations of speakers (Croft, this volume; Kirby, this volume)? What do we know of the history of each particular language? Why are the language faculty and languages the way they are? Why do we see universal tendencies and at the same time great diversity in languages? How does language influence the way we perceive and understand the world? What do we know of the history of the human capacity for speech? Is it mainly the result of genetic evolution, like the evolution of the eyes, or a cultural invention, like writing? Is language an adaptation to a changing environment? An internal change in an individual which increased its chances of reproduction? Is it an exaptation, a side effect of changes which were not at first tied to communicative behaviour? What are the evolutionary prerequisites which paved the way for the capacity of speech? And how did these prerequisites appear? Independently? Genetically? Culturally?

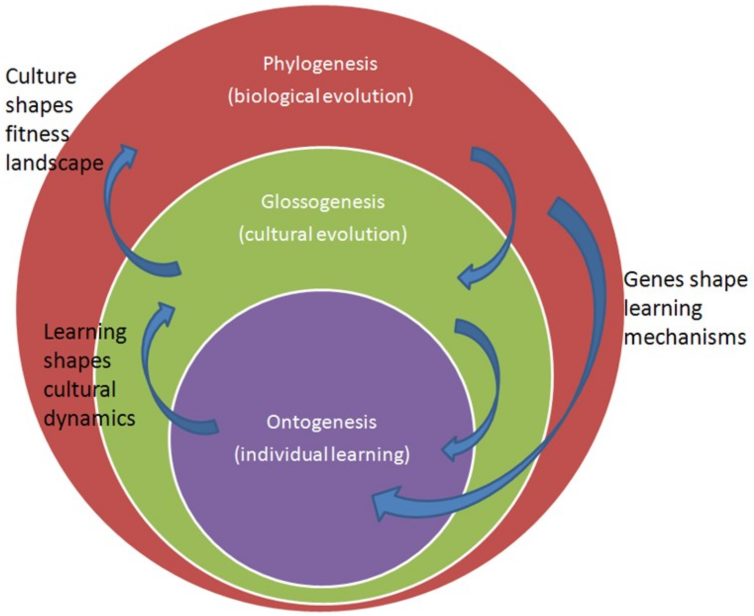


Figure 8. Multiple interaction scales involved in the origins of language

Ranged against the diversity of these questions is an even greater diversity of research disciplines and methods. Linguists, even though they continue to provide crucial data on the history of languages, are no longer the unique actors. Developmental and cognitive psychologists and neuropsychologists carry out behavioural studies of language acquisition and language pathology, and these often reveal cognitive mechanisms involved in language processing. Neuroscience, especially with equipment for brain imaging allowing us to see which brain regions are active for given tasks, attempts to find neural correlates of verbal behaviour, to discover its organization in the brain. Some researchers also study the physiology of the vocal tract, to try to understand how we produce speech sounds. The physiology of the ear, the essential receptor in the speech-decoding chain (or vision, in the case of signed languages), is also a focus of research. Archaeologists examine fossils and artefacts left by the first hominids, and try on the one hand to deduce our anatomical evolution (especially of the larynx) and on the other hand to get an idea of what activities they were engaged in (what tools did they make? how did they use them? What can these tools tell us about the degree of cognitive development?). Anthropologists do fieldwork on isolated peoples, and report on cultural differences, especially those related to languages and the meanings they convey. Primatologists try to report on the communicative capacities of animals that may have some ancestors in common with human and to compare them with our own communicative capacities. Geneticists on the one hand sequence the human genome and that of potential ancestral species when it is possible to specify their phylogenetic relatedness, and on the other hand use genetic information from different people across the planet to help in reconstructing the history of languages, which is often correlated with the genetic history of their speakers (Dediu, this volume).

Thus language involves a multitude of components interacting in complex ways in parallel on several timescales: the ontogenetic timescale, characterizing the growth of an individual person, the glossogenetic or cultural timescale which characterizes the evolution of cultures, and the phylogenetic timescale, which characterizes the evolution of species (see figure 3). In particular, language is characterized by complex physical and functional interactions among multiple cerebral circuits, several organs, the individuals which are equipped with them, and the environment in which they live. Now, as we have seen in previous paragraphs, not only is it essential to study each of these components independently, to reduce the complexity of the problem; it is also necessary to study their interactions. Thus, a growing number of researchers have proposed the idea that many properties of language and languages may not be encoded in any particular component involved, i.e. in certain specific cerebral structures or in properties of the auditory and vocal apparatus, or even in an individual considered independently of others, but rather may be the self-organized outcomes of the complex dynamic interactions among the components and individuals. Yet, these self-organizational phenomena are often complicated to understand or to foresee intuitively, and to formulate in words, whence the crucial use of mathematical and computer modelling which I will now present.

Computer Models and Simulations of the Evolution of Language

Experimenting with Complex Dynamical Systems

Nowadays, one of the most efficient ways to improve our understanding of the dynamics of self-organized systems is the use of computers or robots. Indeed, they allow us to elaborate operational models of which we know all the assumptions, to run them, and to observe the resulting behaviour as a function of the values of parameters set in the context of these models. This is why, in addition to linguists, psychologists, anthropologists, neuroscientists, geneticists, and physiologists, mathematicians and computer scientists/roboticists have now a crucial role in this research.

An operational model is one which defines the set of its assumptions explicitly and above all shows how to calculate their consequences, that is to prove that it leads to a certain set of conclusions. There are two main types of operational model. The first, used by mathematicians and some theoretical biologists, consists in abstracting from the phenomenon of language a certain number of variables along with the rules of their evolution in the form of mathematical equations. Most often this resembles systems of coupled differential equations, and benefits from the framework of dynamic systems theory. The second type, which allows for modelling of more complex phenomena than the first, is that used by researchers in artificial intelligence: it consists in the construction of artificial systems implemented in computers or in robots. These artificial systems are made of programs which most often take the form of artificial software or robotic agents, endowed with artificial brains and bodies. These are then put into interaction with an artificial environment (or a real environment in the case of robots), and their dynamics can be studied. This is what one calls the “method of the artificial” (Steels, 2001) or the “synthetic methodology” (Pfeifer and Scheier, 1999). The use of computational machines to simulate and study natural phenomena is not new: Lorenz used the first computers to study the behaviour of climatological models, Fermi to simulate non-linear interactions among magnetized particles, Turing to imagine how morphogenetic processes could self-organize, Von Neumann to study self-replication.

More recently, this method has allowed ethology to progress significantly in the understanding of behaviour performances of social insects (Bonabeau et al., 1997). Computer simulations of social insects, based on the concept of software or robotic agents modelling each insect individually, hence the term agent-based modelling, were built. This has permitted to establish sufficient characteristics of behaviour and insect capabilities that lead to the formation of collective structures, such as the construction of termite nests, or the formation of organized groups for hunting or foraging in ants, or the formation of fish shoals, thermoregulation in beehives or the formation of social structures in wasps. In general, these computer simulations have shown that it was often not necessary that insect be equipped with complex cognitive structures so that we can observe the collective formation of complex structures.

Physicists have also a tradition of using computers to build simulations of complex systems that allow them to elaborate their intuitions. For example, through the experimentation with cellular automata – sorts of grids which cells can be in an “on” or “off” state and evolution depends on the state of their neighbours according to simple rules – they discovered how, starting either from initially random structures or completely uniform structures, complex patterns with non-trivial symmetries could be formed, resembling those observed in ice crystals, in the distribution of avalanches in sand piles or in mountains, dunes in the desert, the shape of fluvial deltas, galaxies or polyhedral bubbles in water cascades. For physicists, cellular automata are not what could be called physical models of ice crystals or avalanches, but they have played the role of metaphors and analogies which triggered a renewal in the way their community perceived and understood these phenomena (Vichniak et al., 1989; Weisbuch, 1991; Bak, 1996; Ball, 2001).

Computer Science and the Origins of Language and Languages

It is also possible to use computers and agent-based simulations not only to help us understand the phenomena that characterize self-organization of matter, simple biological structures, or insect societies, but also to help us understand phenomena that characterize humans and its societies. The time has come to use computers and robots as scientific tools in human sciences. Thus, building artificial systems in the context of research into language origins and the evolution of languages is enjoying a growing popularity in the scientific community, exactly because it is a crucial tool for studying the phenomena of language in relation to the complex interactions of its components (Steels, 1997; Oudeyer and Kaplan, 2007; Kaplan and Oudeyer, 2008). These systems are put to two main types of use: 1) they serve to evaluate the internal coherence of verbally expressed theories already proposed by clarifying all their hypotheses and verifying that they do indeed lead to the proposed conclusions (and quite often one discovers errors in the assumptions as well as in the conclusions, which need to be revised); 2) they serve to explore and generate new theories, which themselves often appear when one simply tries to build an artificial system reproducing the verbal behaviour of humans.

A number of decisive results have already been obtained and have opened the way for resolution of previously unanswered questions: the decentralized generation of lexical and semantic conventions in populations of agents (e.g. Kaplan, 2001), the formation of shared inventories of vowels or syllables in groups of agents (e.g. Berrah et al., 1996; de Boer, 2001; Oudeyer, 2001; Oudeyer, 2005a; Oudeyer, 2005b; Oudeyer, 2006), with features of structural regularities greatly resembling those of human languages (e.g. Pierrehumbert, 2001; Wedel, 2006), the formation of conventionalized syntactic (e.g. Batali, 1998) and grammatical structures (e.g. Steels, 2005), the conditions under which combinatoriality, the property of systematic re-use, can be selected (Kirby, 2001).

It is crucial to note that in the context of research on the origins of language, this methodology of the artificial is a *methodology for exploration*. It fits within an abductive scientific logic, i.e. a logic in which one is searching for the premises that can lead to a given conclusion (instead of a deductive logic in which one searches for the conclusions that can be reached from a given set of premises).

The word ***model*** has here a different meaning than in its traditional use. Indeed, traditionally, modelling consists in observing a natural phenomenon and then to abstract fundamental mechanisms and variables on which a formalism is constructed that permits to predict reality as precisely as possible. Here, we are rather interested in qualitative investigations of broad types of mechanisms that may have been implemented in nature to solve given problems. Language is such a complex phenomenon that observations alone cannot allow researchers to *deduce* explanatory mechanisms. On the contrary, it is necessary to have beforehand a structured and rich conceptualisation of the space of hypotheses and mechanisms that might explain the complex properties of language evolution. This is where artificial computational systems, for which we use the term ***model,*** come into play: they are used to refine our intuitions on the dynamics of language and languages formation, as well as to sculpt the space of hypotheses.

As a consequence, the aim of these computational models is not to establish the list of mechanisms responsible for the origins of given aspects of language. Rather, the objective is more modestly to try to build a list of potential candidates, to constrain the space of hypotheses, in particular by showing examples of mechanisms that are sufficient and examples of mechanisms that are not necessary.

The Speech Code

I will now illustrate this computer modelling work about the evolution of language and languages with the description of an experiment which focuses on the problem of the origins of speech, i.e. sound systems as physical vehicles of language (as gestures can also be in signed languages). The goal of this experimentation is to foster the reconceptualisation of this scientific issue, through the evaluation of both existing and novel scientific hypotheses.

Discreteness and combinatoriality.

Humans have a complex system of vocalizations. They are discrete and combinatorial, i.e. they are built from elementary units, "sculpted" in the auditory and phonatory continuum, which are systematically recombined and reused. These units exist at several levels (motor primitives to obstruct the air flow in the vocal tract, called *gestures*; gestures coordinations, called phonemes and which define vowels and consonants; syllables; etc …). Whereas the articulatory space is continuous and potentially permits an infinity of gestures and phonemes, each language discretizes this space in its own way, carving a repertoire of gestures and phonemes both small and finite (Studdert-Kennedy and Goldstein, 2003). This is why it is sometimes referred as *phonemic coding*.

Universals and diversity.

In spite of the great diversity of these elementary units in world languages, one can also find strong statistical regularities. For example, certain vowel systems are much more frequent than some others, such as the five-vowel system composed of vowels [e], [i], [o], [a] and [u]. The same fact can be stated for consonants. The way units are combined is also peculiar: one the one hand, not all sequences of phonemes are allowed to form syllables in each particular language, and on the other hand the associated sets of possible phoneme combinations can be grouped into generic types. This organization into generic types means that for example, one can summarize the possible phoneme combination to form syllables in Japanese ("moras") with types "CV/CVC/VC", where "VC" denotes syllables composed of two slots, with any Japanese vowel in the first slot and any Japanese consonant in the second slot.

Cultural sharing.

Speech is a conventional code. Whereas statistical regularities can be observed across human languages, each linguistic community possess its own way to perceive and categorize sounds as well as its own repertoire of rules of combinations. For example, native speakers of Japanese do not hear the difference between the *r* in *read* and the *l* in *lead*. How can a linguistic community come to form a code shared by all individuals and without a central coordinated control of the code?

Since the work of de Boer (2001) and Kaplan (2001), we have convincing hypotheses about how a new sound or a new word can propagate and be accepted in a given population. But these negotiation mechanisms, also called "consensus dynamics", assume the pre-existence of linguistic interaction conventions (Croft, to appear). Thus, the associated models concern rather the formation and evolution of languages, but do not address directly the question of the origins of language. Indeed, when there was not already a conventional linguistic communication system, how could the first conventions have bootstrapped?

The model I will present focuses on this later question. It is obviously linked to the question of the origins of languages, because it is about understanding how a speech code may have formed to be used as a basis for the first languages. The main difference among the two questions lies in the properties that shall characterize the mechanisms we are searching for. If one is interested in the origins of speech, one must search for an explanatory mechanism which assumes neither the existence of linguistic conventions, nor the existence of cognitive structures that are specific to language. Indeed, this would imply that we would have models of individuals that can already speak, and thus for which language would already have appeared.

Self-Organization and the Evolution of Speech

How did human first speech codes formed in a society with no language? As argumented above, two kinds of answers must be brought. First, a functional answer: it establishes the function of vocalization systems, and shows that human systems, with the properties we described, are efficient to achieve this function. Liljencrantz and Lindblom (1972) proposed such an answer, and showed that the statistical regularities of human repertoires of phonemes were the most efficient in terms of the perceptual distinctiveness/articulatory cost compromise. This kind of answer is necessary, but not sufficient: it does not allow us to explain how evolution (genetic or cultural) could have found those quasi-optimal structures, and does not allow us either to explain how a given linguistic community can "choose" one solution among several quasi-optimal ones. In particular, it is possible that "naïve" Darwinian search with random mutations is not efficient enough to find complex structures such as those of the speech code: the search space is too large.

This is why a second kind of answer is necessary: we need to investigate how biological evolution might have generated and selected these structures. One possibility is to study how self-organization may have constrained the search space to help natural selection. This would consist in showing that a system much simpler than the structure we want to explain spontaneously *self-organizes into this structure*.

I will now present such a system and show how relatively simple premises – from an evolutionary point of view – can lead to the self-organized formation of speech codes.

A Computer Investigation of the Formation of Fundamental Structures of Speech

This computer model is agent-based: it consists in setting up virtual robots equipped with models of the auditory and phonatory apparatus coupled with a network of artificial neurons that connect perceptual and motor modalities. These artificial neurons determine the robots' behaviour, mainly consisting in vocal babbling. The babbling activity, coupled with the properties of plasticity characterizing neural networks, allow the robots to learn the correspondences between the space of auditory perceptions and the space of vocal tract gestures. Finally, these robots are placed together in a shared environment where they can hear the vocalizations of their neighbours, which influence their own vocalizations, and wander around. I will show that a number of emergent properties characterizing the vocalizations produced by robots in a given population form spontaneously.

More technically[25], agents possess an artificial ear (which properties can be modified to study their specific role, see below) capable of transforming an acoustic signal into neural impulses which stimulate neurons in an artificial perceptual neural map. They also possess a motor neural map which neurons activation produce movements of a vocal tract model, which itself produces an acoustic wave (and which degree of realism can equally be modified). Both perceptual and motor maps are totally interconnected (see figure 4). Initially, internal parameters of all neurons, as well as the parameters of their connections, are random. To produce a vocalization, a robot randomly activates several motor neurons, which internal parameters encode articulatory configurations which shall be reached in sequence, which then produces an articulatory trajectory and, through the vocal tract model, an acoustic signal that can be perceived by

---

[25] We only give here a general description of the system: a precise mathematical description is available in (Oudeyer, 2006).

the ear model. This is the basis of babbling, and explains why initially, robots produce vocalizations randomly spread in the vocal continuum. These neural networks are characterized by two forms of plasticity: 1) intermodal connections evolve in such a way that the robot learns the correspondences between auditory and motor trajectories perceived and produced when it is babbling[26]; 2) neurons in each map evolve in such a way that they tend to model the distribution of sounds heard by the robot[27]; 3) The connections between both neural maps are such that the distribution of sounds encoded in the motor map follows roughly the distribution of sounds encoded in the perceptual map. This implies that the neural architecture is such that robots have the tendency to produce the same distribution of sounds as the distribution of sounds that they hear around them. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself on the distribution of sounds in this language. For example, is this language contains the vowels [a,e,i] but not [o], the robot's babbling vocalizations will quickly contain [a,e,i] much more often than [o]. This behaviour corresponds to what is observed in young infants, and referred as "phonological attunement" (Vihman, 1996).

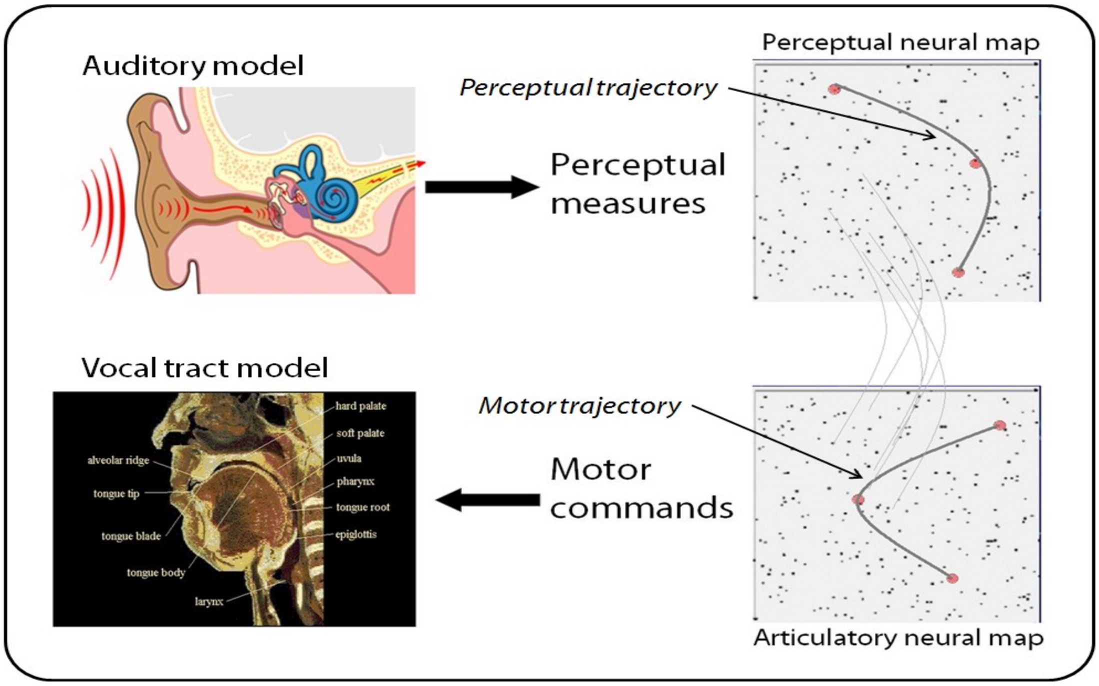


Figure 4. The robot architecture in the artificial system. Each robot is equipped with an artificial ear, an artificial vocal tract, and two neural maps (perceptual and motor) that couple the ear and the vocal tract. These neural maps are initially random but characterized by two forms of plasticity: 1) intermodal connections evolve in such a way that the robot can learn the correspondences between auditory trajectories and associated motor trajectories when the robot is babbling; 2) neurons in each map evolve in such a way that they track the distribution of sounds heard by the robots. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself on the distribution of sounds in this language.

---

[26] Connections between the two maps evolve according to Hebb's law: those that link neurons that are often activated in a correlated manner are reinforced, whereas those that link neurons with uncorrelated activation become weaker. These connections are initially random, and through babbling and Hebb's law, they self-organize and finally allow the robot to find motor commands that correspond to a given sound that he perceives.

[27] Neurons adapt to stimuli through sensitization: their dynamics is such that if a stimulus S is perceived, then they are modified such that if the same stimulus S would be presented just after they would be more activated than the first time, and the amount of modification depends exponentially on their activation (strongly activated neurons are modified most).

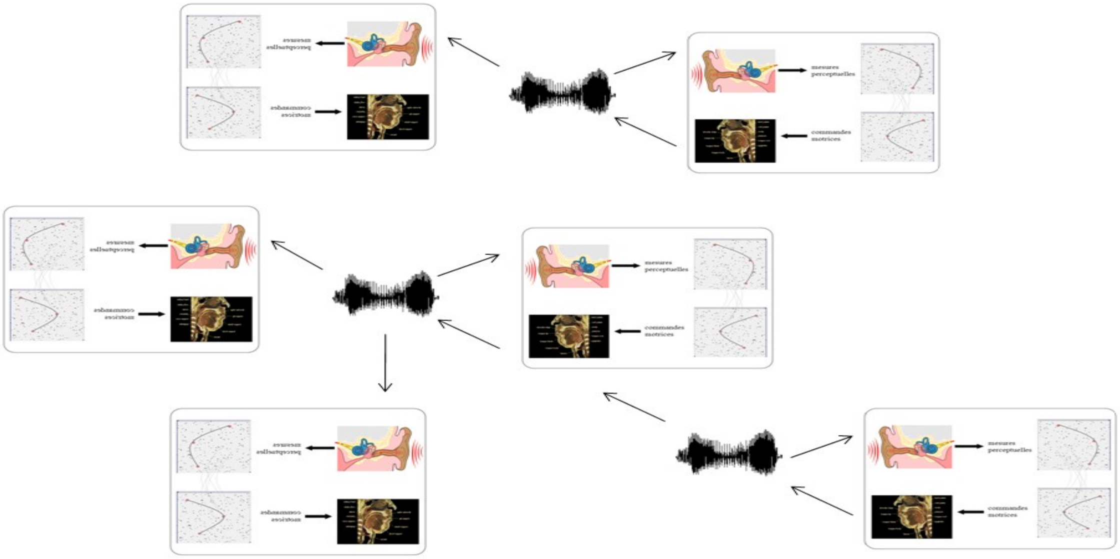


Figure 5. In the experiment, babbling robots are placed in a shared environment and can hear each other's vocalizations in addition to their own vocalizations. Their neural networks are initially random, thus their vocalizations are initially random and unorganized. Given that the properties of plasticity of their brains push them to align their vocalizations with the one they hear, and as they statistically all produce vocalizations randomly uniformly spread in the articulatory space, the initial state is an equilibrium. Yet, this equilibrium is unstable and the noise inherent to their interactions and to the neural dynamics provokes a symmetry breaking: vocalizations crystallize and become discrete, combinatorial and shared by all the members of the population.

A Unified Mechanism for the Self-Organization of Combinatoriality, of the Universals/Diversity Duality, and of Cultural Sharing

This type of architecture has frequently been used in the literature to model speech acquisition in children (Kohonen, 1988, Sanguineti et al., 1998), in experiments in which the system learnt to pronounce sounds/syllables of a language it was exposed to. Yet, the experiment I present here is different: one does not assume that a constituted speech code exists initially. On the contrary, one places a population of babbling robots together in a shared environment, such that they can both perceive their own babblings and those of their neighbours (see figure 5). Given that the properties of plasticity of their brains make them align their babbling vocalizations on those they hear around them, and as initially they all produce random vocalizations uniformly spread in the vocal space, the initial state is equilibrium.

Yet, if one runs the simulation, one observes that this equilibrium is not stable. Indeed, noise – stochasticity – makes that from time to time, certain types of vocalizations are pronounced a little more often than others. Now, the mechanism that couples perception and production creates a positive feedback loop: these deviations from the mean get amplified when they are big enough, and the system's symmetry breaks. Neural maps self-organize into clusters of neurons encoding particular acoustic and articulatory configurations in the space of vocalizations (see figure 6). In brief, the continuous space of vocalization has been discretized. The vocalizations produced by agents are not holistic anymore, but discrete and combinatorial: they are systematically built through the sequencing of key configurations, that we can call phonemes. One sees the formation of phonemic coding, i.e. discrete combinatorial speech codes as described earlier. Besides, the system of phonemes that self-organize is shared by all robots of a given simulation, and is different in different simulations. Thus, one observes the formation of a "cultural convention" which can be diverse across groups.

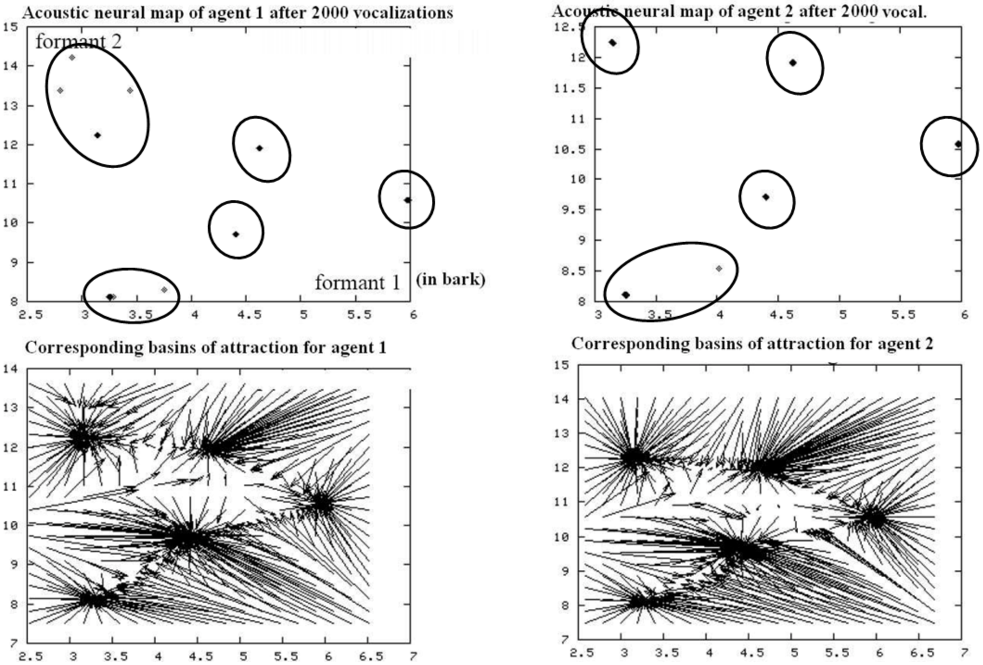


Figure 6. Very quickly, initial symmetry in the system is broken, and neurons which initially encoded random vocal configurations now encode a small number of configurations which are systematically reused by agents when they babble: the vocal space has been discretized. Besides, these self-organized elementary configurations are the same in all robots of the same population, but different across populations. One can see it on this figure which represents the perceptual neural maps of two agents after 2000 vocalizations, where the configurations encoded by neurons are clustered (top), as well as their corresponding distributions (bottom). The auditory space is here projected on the first and the second effective formant, expressed in barks, which allows us to visualize the self-organized vowel systems.

As a matter of fact, several variants of this experiment can be set up and permit to refine the conclusions. First, it is possible to experiment what happens when there is only one single robot which is listening to itself babbling. In this case, one also observes a crystallization of vocalizations: it quickly ends up producing vocal trajectories that systematically reuse few key articulatory configurations. One can deduce that the formation of phonemic coding, i.e. discreteness and combinatoriality, is not the result of social interactions but rather of the internal coupling between speech perception and production. Yet, whereas the vocalizations of isolated babbling agents will crystallize on different vocalization systems, these systems will spontaneously synchronize when they share the same environment and are capable of hearing each other: in this case, the self-organized systems are approximately the same in all agents on the same population.

A second important variant of this experiment consists in varying the morpho-physiological properties of the auditory and phonatory systems so as to determine the impact of these properties on the systems that form (or do not form). In particular, a crucial property of the speech organs is the non-linearity of the function that maps articulatory configurations to acoustic waves to auditory perceptions. The human vocal tract is indeed such that for certain articulatory configurations, small variations produce small variations in the perceived sound, while for other articulatory configurations, small variations produce large variations in the perceived sound. Now, this property is central in several theories that propose to explain why speech is phonemically coded, such as in Stevens' quantal theory (Stevens, 1989) or in the DRM model (Mrayati et al., 1988). It is possible to use a model of the ear and the vocal tract that are realistic and include this type of non-linearity, but it is also possible to construct on purpose a non-realistic model to evaluate the specific impact of the non-linearities. These experiments were run, and with a linear model, one observes that in a population of babbling robots the crystallisation we presented above still happens: vocalizations self-organize into a combinatorial system in which particular articulatory configurations are systematically reused in vocal trajectories. Thus, we can make a first conclusion: these simulations show that phonemic coding can appear spontaneously without non-linearities in the auditory-phonatory system. This does not imply that non-linearities do not accelerate the formation of phonemic coding, but that they are not necessary, as proposed in the quantal theory or in the DRM model.

Yet, if one looks at the distribution of self-organized key vocal configuration when one uses the linear auditory-phonatory system (which one can see as kinds of phonemes), one observes that these configurations are globally positioned randomly uniformly in the space of possible vocal configurations. But when one uses a realistic ear and vocal tract model, reproducing in particular the properties of vowel production and perception[28], one observes an additional phenomenon. Besides the crystallization which is the qualitatively same as with the linear model, the vocalization systems that form are characterized by statistical regularities that share many similarities with human vowel systems. For example, one can collect statistics on the vowel systems that appear as key reused configurations in the self-organized systems when one runs many simulations. The results, illustrated in figure 7, show that on the one hand a diversity of systems appear, and on the other hand that certain vowel systems appear much more often than others. As a conclusion, one observes the same duality of universals and diversity that characterized human languages and the simulation proposes a unified explanation:

1. The dynamical system composed by the set of babbling robots and the internal sensorimotor couplings possess a number of attractors which are culturally shared combinatorial vocalization systems;
2. Under the influence of noise and small variations of initial conditions, the dynamical system falls in a particular attractor, which allows us to explain the “decentralized collective choice” made by the population to adopt a system rather than another;
3. Non-linearities in the auditory and phonatory systems introduce asymmetries between attractors: some of them have a larger basin of attraction, in particular those for which phonemes are in zones where small articulatory variations provoke small perceptual variations, which increases the probability that the system falls in such attractors;

Furthermore, there is not only a structural correspondence between simulations and reality, but the vowel systems that appear most frequently in robot populations are approximately the same, and in the same proportion, than those that appear most frequently in human languages. Thus, there is a quantitative relationship. One can conclude that the non-linearities of the auditory and phonatory systems are decisive to explain why certain systems of phonemes are statistically more frequent than others. Yet, the very existence of these phonemes, i.e. the existence of a vocalization system in which invariant articulatory and auditory configurations are systematically reused, is not necessitated by these non-linearities.

---

[28] See (Oudeyer, 2006) for a precise description of the model based on the work of (de Boer, 2001).

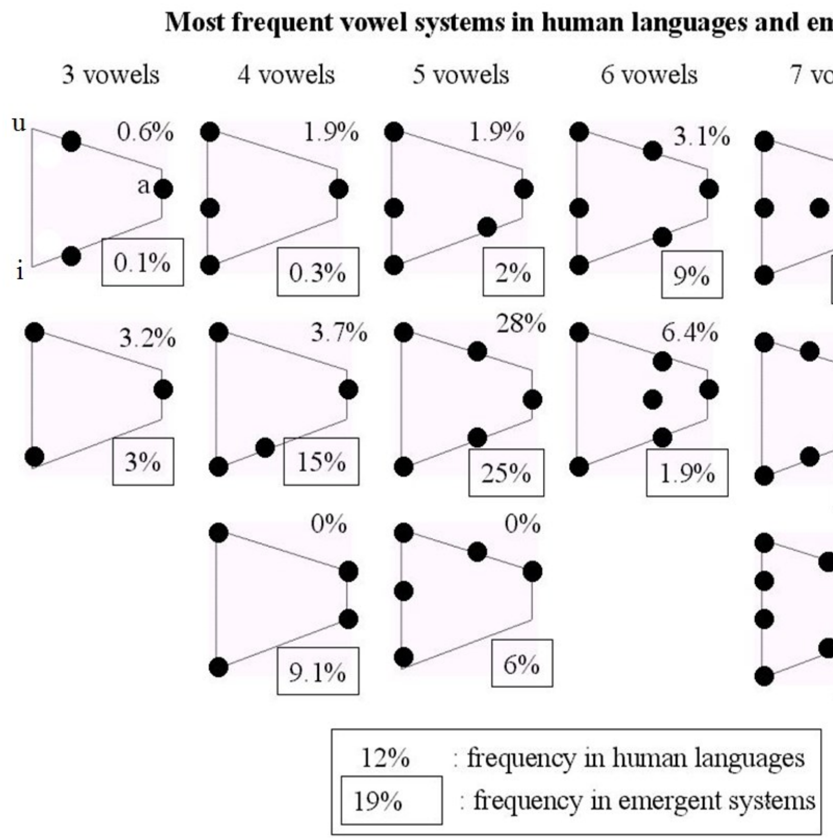


Figure 9. Comparison between the distribution of vowel systems formed by populations of robots and in human languages of the UPSID database (Maddieson, 1984). Vowel systems are represented on the vocalic triangle, which horizontal dimension corresponds to the first formant, and vertical dimension corresponds to the second effective formant. One observes that the most frequent systems in the artificial system are the same as in human languages (in particular the symmetric 5 vowel system /a,e,i,o,u/ with 25 percent in artificial systems and 28 percent in human languages).

It is important to note that in these simulations, neural architectures are characterized by several parameters, and that all values of these parameters do not lead to the above mentioned results. Nevertheless, only one parameter has a crucial influence on the results (Oudeyer, 2006): neurons are characterized by a selectivity to stimuli (σ) that can be focused or wide. If this selectivity is too focused, no crystallization happens, but the robots remain capable to learn the relationships between the auditory space and the phonatory/motor space. If the selectivity is too wide, the system crystallizes in a degenerate state in which all vocalization are exactly the same and are stationary: there is only one phoneme. Yet, there is a large range of intermediary values between these two extremes that allows us to observe a crystallization in which a combinatorial system with multiple phonemes self-organizes in a population of babbling robots.

Towards a Novel Vision of Evolutionary Scenarios of the Evolution of Speech.

The preceding discussion has important consequences if one uses the model to imagine evolutionary scenarios that may have led to the formation of the first vocalization systems featuring the fundamental properties of contemporary human speech. Indeed, they imply that many parameter variations in the neural architecture permit to generate combinatorial speech systems shared by all members of a community. Also, they imply that with such a neural architecture, a combinatorial phonemically coded vocalization system can appear without assuming special properties of the auditory and phonatory system: the only assumption is that a certain variety of sounds can be produced, but non-linearities are not required. Finally, the architecture is in itself relatively primitive: it includes neural units which intrinsic properties (dynamics and plasticity) are classical and are functionally similar to most of neural units in mammal brains (Oudeyer, 2006). The specificity of this architecture relies in the systematic and direct plastic connections between the auditory and motor maps. Yet, this specificity characterizes the basic building blocks of the capacity to learn to imitate holistic sounds, which we call here ***holistic adaptive vocal imitation*** and is sometimes also called adaptive vocal mimicry or vocal learning in the animal literature, which is a capacity functionally more restricted than the capacity to learn and share combinatorial phonemically coded speech sounds. This leads us to the following evolutionary scenario to conceptualize the origins of combinatorial vocalization systems culturally shared by all individuals of a community:

1) Adaptive vocal imitation is present in many animals (Snowdown and Hausberger, 1997; Hauser, 1997) which possess shared and learnt vocalization systems, but which do not possess language. Ethologists have identified many potential reproductive advantages characterizing the capacity of adaptive vocal imitation in a community of individuals (i.e. this allows individuals to mark their group membership). Thus, it is reasonable to think that before being capable of speech and language, humans may have evolved the capacity to imitate vocally;
2) Being capable of adaptive vocal imitation, as well as the related reproductive advantages identified in non-human animals, does not imply and does not necessitate a combinatorial phonemically coded vocalization system. As an example, the range of parameter values for which the selectivity σ is focused allows the robots to learn efficiently the vocal perceptuo-motor correspondences without generating a phonemic system;
3) Now, if one imagines an ecological context in which the presence of a combinatorial vocalization system would provide a reproductive advantage to those that possess it, then the experiments that we described permit to state that a simple change in the value of the σ parameter in motor and perceptual neural map would lead to the spontaneous formation of vocalization systems which feature several fundamental properties of contemporary human speech systems. This allows us to understand that what may have been a great leap forward for language, i.e. the formation of shared combinatorial vocalizations, may be the consequence of a small biological change together with the self-organizing properties of neural matter and multi-modal sensorimotor coupling.

Moreover, this scenario in which phonemically coded vocalization systems would have been selected thanks to the reproductive advantage that they may have provided is not the only one that the computational model can support and refine. Indeed, I explained above that in the range of σ values that allows combinatorial systems to be formed, the capacity of adaptive vocal imitation is intact and equally efficient. Besides, performance being equal, the transition of σ among this range and the range of more focused values for selectivity does not a priori imply a metabolic cost. This implies that in an ecological context in which those neural structures appeared under a selective pressure for adaptive vocal imitation, neutral mutation/variation and neutral drift may have happened and generated spontaneously shared phonemically coded vocalization systems without a selective pressure for language. One observation makes this scenario particularly stimulating: among animal species capable of adaptive vocal imitation in which culturally shared sound systems exist, but which do not possess language, several of them produce vocalizations or songs structured around the systematic reuse of basic units. For example, this can be observed in zebra finches (Brenowitz and Beecher, 2005) or in humpback whales (Tyack, 1981). The function of this quasi-phonemic structuring has been only little understood so far in ethology. Besides, because the model I presented is neutral with respect to many properties of the auditory and phonatory systems, and because the neural architecture that it assumes corresponds to the minimal equipment for adaptive vocal imitation, it can be applied to the formation of songs in these animals. In this case, it provides a hypothesis reinforced by the current scientific uncertainty about the function of combinatorial coding in these songs: combinatorial and systematically reused units may have been generated spontaneously as a collateral effect of the biological equipment for adaptive vocal imitation. Thus, it is also reasonable to imagine that this may have been the case in humans: combinatorial speech systems may have been recruited only later on to achieve their current linguistic function. This implies that several fundamental properties of contemporary human speech systems may be exaptations[29].

Conclusion

Through the elaboration and use of a computer model, I showed how a relatively simple neural architecture coupling auditory and phonatory modalities permitted, via a self-organizing dynamics, the spontaneous formation of combinatorial phonemically coded vocalization systems shared by all members of a community, and characterized by the duality of universals and diversity. The first contribution of this work is that for the first time it provides a unified explanation of these three phenomena.

Besides, this multimodal coupling architecture corresponds to the minimal neural kit required by adaptive vocal imitation, and does not include biological elements that are specific to human speech. Given that crystallization happens in a large range of the parameter space, this shows that the transition from inarticulated vocalization systems to human-like speech codes may have been largely due to a modest biological innovation. Indeed, the model indicates that neuronal structures that encode a priori and specifically phonemic organization, as well as typical regularities of speech, do not need to be innately generated to allow the formation of such speech code. This is the second contribution of this work: it allows us to understand how the self-organizing properties of simple neural structures may have constrained the space of biological vocalization structures and how speech codes may have been generated and selected during phylogenesis.

These new hypothesis may not have been identified without the use of computer models and simulations, because the underlying dynamics are complex and difficult to anticipate through uniquely verbal reasoning. This illustrates the potential importance of these new methodological tools in human and biological sciences. Yet, these computer models abstract many biological and behavioural mechanisms, and consist primarily of a theoretical investigation of the space of hypotheses: once this space is reconceptualised, and the internal coherence of hypotheses evaluated through computer simulations, validation work and grounding of these hypotheses in biological field observations remains to be done. Thus, the third contribution of this work is, more than the elaboration of specific hypotheses, the construction of a framework and of tools that allow us to develop new intuitions and new concepts for our understanding of the origins and evolution of speech.

Acknowledgments: This work was in major part achieved in the Sony Computer Science Laboratory, Paris, and benefited from the support of Luc Steels.

References

Ashby, W.R. (1956) *An Introduction to Cybernetics*, Chapman & Hall.

Bak, P. (1996) How Nature Works: The Science of Self-Organized Criticality, New York: Copernicus.

Ball, P. (2001) The self-made tapestry: pattern formation in nature, Oxford: Oxford University Press.

Batali J. (1998) 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 University Press.

---

[29] This term was introduced in (Gould and Vrba, 1982). It refers to the use of a biological feature/structure for a function A which is different than the function B for which it was initially evolutionary selected.

Berrah, A., Glotin, H., Laboissière, R., Bessière, P.& Boë, L.J. (1996) From form to formation of phonetic structures: an evolutionary computing perspective, in: 13th int. conf. on Evolutionary computing & Machine Learning (ICML), pp23-9, MIT Press, Bari-Italy.

Bonabeau, E., Theraulaz, G., Deneubourg, J.L., Aron, S. & Camazine, S. (1997) Self-organization in social insects. *Trends in Ecology and Evolution*, 12: 188-193.

Brenowitz, E.A. & Beecher, M.D. (2005) Song learning in birds: diversity and plasticity, opportunities and challenges, *Trends in Neuroscience, 28(3)*, pp. 127—132.

Camazine, S., Deneubourg, J-L., Franks, N.R., Sneyd, J., Theraulaz, G. & Bonabeau E. (2002) *Self-Organization in Biological Systems*, Princeton University Press.

Croft W. (to appear) Language use and the evolution of languages, in *The Language Phenomenon*, eds. K. Smith and P. Binder, Springer.

Dawkins, R. (1986). The Blind Watchmaker. New York: W. W. Norton & Company.

de Boer, B. (2001) *The origins of vowel systems*, Oxford Linguistics, Oxford University Press.

Dediu, D. (to appear) Genes and language: inter-individual variation, historical correlations and genetic biasing, in *The Language Phenomenon*, eds. K. Smith and P. Binder, Springer.

Elredge, N. and Gould S.J. (1972) Punctuated equilibria: an alternative to phylogenetic gradualism, in T.J.M. Schopf (ed.), *Models in Paleobioogy*, pp. 82—115, San Francisco: Freeman.

Gould, S. J. & Vrba, E.S. (1982) Exaptation — a missing term in the science of form, *Paleobiology,* 8 (1): 4–15

Hauser, M. D. (1997) *The Evolution of Communication*. MIT Press/BradfordBooks.

Kauffman, S. (1996) At home in the universe: the search for laws of self-organization and complexity, Oxford: Oxford University Press.

Keefe, A. & Szostak, J. (2001) Functional proteins from a random sequence library, *Nature*, 410, pp. 715—18.

Kaplan, F. (2001) La naissance d’ une langue chez les robots; Hermès.

Kaplan, F., Oudeyer, P-Y. & Bergen B. (2008) Computational Models in the Debate over Language Learnability, *Infant and Child Development*, 17(1), pp. 55—80.

Kirby, S. (2001) 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.

Kirby, S. (to appear) The evolution of linguistic replicators, in *The Language Phenomenon*, eds. K. Smith and P. Binder, Springer.

Kohonen, T. (1988) The neural phonetic typewriter, *Computer*, 21(3), pp. 11—22.

Liljencrantz, J. & Lindblom, B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language,* 48: 839-862.

Maddieson, I. (1984) *Patterns of sounds*, Cambridge: Cambridge University Press.

Mrayati, M., Carre, R. & Guerin, B. (1988) Distinctive regions and modes: a new theory of speech production, *Speech Communication*, 7, pp. 257—286.

Nicolis, G. & Prigogine, I. (1977) Self-organization in non-equilibrium systems: from dissipative structures to order through fluctuations, New-York: Wiley.

Oudeyer, P.-Y. (2001) Origins and learnability of syllable systems, a cultural evolutionary model, in P. Collet, C. Fonlupt, J. H. E. L. M. S. (Ed.), *Artificial Evolution*; LNCS 2310; pp. 143–155.

Oudeyer, P-Y. (2005a) The Self-Organization of Speech Sounds, *Journal of Theoretical Biology*, 233(3), pp. 435--449.

Oudeyer, P-Y. (2005b) The self-organization of combinatoriality and phonotactics in vocalization systems, *Connection Science*, 17(3-4), pp. 325—341.

Oudeyer, P-Y. (2006) *Self-organization in the evolution of speech*, Oxford University Press.

Oudeyer, P-Y. & Kaplan, F. (2007) Language Evolution as a Darwinian Process: Computational Studies, *Cognitive Processing*, 8(1), pp. 21--35.

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. and P. Hopper (Eds.) *Frequency effects and the emergence of linguistic structure*. John Benjamins, Amsterdam.

Pfeifer R. & Scheier C. (1999) *Understanding Intelligence*, Cambridge: MIT Press.

Sanguineti, V., Laboissière, R. & Ostry, D. J. (1998) A Dynamic Biomechanical Model for Neural Control of Speech Production. *Journal of the Acoustical Society of America*, 103(3):1615-1627.

Snowdown, C.T. & Hausberger, M. (1997) *Social influences on vocal development*, Cambridge University Press.

Steels, L. (2001) The Methodology of the Artificial, *Behavioral and Brain Sciences*, 24(6).

Steels, L. (2005) The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems. Connection Science, 17(3-4):213-230.

Stevens, K.N. (1989) On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.

Studdert-Kennedy, M .& Goldstein, L. (2003) Launching language: the gestural origin of discrete infinity, in Christiansen, M. and Kirby, S. (eds.), *Language evolution: the state of the art*, pp. 235—254, Oxford: Oxford University Press.

Tritton, D.J. (1988) *Physical fluid dynamics*, Oxford: Oxford University Press.

Tyack, P. (1981) Interactions between singing hawaian humpback whales and conspecifics nearby, *Behavioral ecology and sociobiology*, 8(2), pp. 105—116.

Vichniac, G., Manneville, P., Boccara, N. & Bidaux R. (Eds.) (1989) *Cellular automata and modeling of complex systems*, in: Workshop Les Houches, Springer, Heidelberg.

Vihman, M. (1996) Phonological development: the origins of language in the child, Cambridge, Mass.: Blackwell.

Waldrop, M. (1990) Spontaneous order, evolution, and life, *Science*, 247, pp. 1543—5.

Weisbuch, G. (1991) *Complex systems dynamics*, Santa Fe Institute Studies in the Sciences of Complexity Lecture Notes, Vol. 2.

Wedel, A. (2006) Exemplar models, evolution and language change. *The Linguistic Review*, 23: 247-274.

Report series Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS)

Editors: Lacerda, Francisco; Bjursäter, Ulla

PERILUS 2011: Lacerda, Francisco

Symposium on Language Acquisition and Language Evolution

www.ling.su.se/perilus-2011

ISSN 0282-6690