

Languages as Memetic Organisms

Gréagóir Bogan

We can make useful deductions about natural languages by modeling them as (clades of) memetic organisms in a competitive environment.

These organisms are subject to evolutionary pressures similar to those applied to all biological organisms, so we can take lessons from evolutionary biology and apply them to languages.

The Organisms

There are two levels at which we might choose to examine the memetic-reproductive properties of natural languages.

The first is considering what we normally consider a “Language” (English, Japanese, etc.) as an individual organism. This model doesn’t work very well, and it doesn’t really allow us to apply any lessons from population genetics.

The second model, which I find vastly more useful, is to consider Languages as *clades* of individual organisms. I will call the individual organisms “*Linguomes*” (in the pattern of genome, memeome, etc.). A Linguome is the set of linguistic patterns reified *in an individual person’s brain*.

A language is a *clade* of closely related Linguomes. When we say “Japanese”, we refer to an abstract categorization which picks out a set of structurally similar linguomes (or to the selected linguomes themselves).

It is widely understood that the English spoken by one person is not exactly the same as the English spoken by another; they will have slight differences in vocabulary, idiom, etc., but their linguomes are so similar that we consider them to be members of the same species.

We will consider the reproductive characteristics of languages in more detail.

The Nature of the Environment

Every person has a *Linguome*. This is comprised of their syntactic models, vocabulary, idioms, etc. One might consider that a polyglot has either one linguome or multiple linguomes; either model works fine, although saying that

a person who speaks both English and Spanish is bilinguomic is probably directionally more accurate.

Linguomic reproduction

Languages have a reproductive strategy which is not common in biological organisms.

In particular, gene transfer is *exclusively* horizontal, similar to bacterial horizontal gene transfer via plasmids. There is no sexual reproduction *or* parthenogenesis. Continuing in the pattern of “gene:genome”, we will use “ling:linguome”. A “ling” is a transmissible packet of linguistic data, such as a word, phrase, or syntactic rule.

Whenever a new human is created, it has some linguistic capacity which it seeks to fill (and which can be stretched to varying degrees beyond satiety, depending on the individual). The new human will populate its empty linguomic capacity with lings (syntactic rules, word definitions, etc.) acquired from nearby linguomes (parents, family friends, other children, media, etc.).

Linguomic transfer continues into adulthood, with individuals absorbing lings from schooling (either of their primary language or of secondary languages), from peers (slang), and in a professional context (field-specific jargon).

Environmental Pressures

As in the genetic evolutionary environment, there are various selective pressures acting on languages.

Brain Power

There is a finite supply of cognitive capacity available for languages to occupy. Languages compete for cognitive capacity in the same manner as memes.

Every individual carrying a language dedicates some amount of capacity to near-universal core concepts of the language (grammatical structure, common vocabulary, etc.). These may be analogized to the set of *fixed* alleles (the versions of various genes which are essentially universal in a population, because they are so essential that any deviation renders them non-viable or incompatible with the broader population).

Individuals also have some capacity for non-universal lings, such as domain-specific jargon, advanced syntactic and grammatical constructs, knowledge of historical or formal variations of lings, etc.

We can quantify the complexity of a language by looking at the population-wide “genetic variance”, so to speak, of the population. We might call this the “metalinguomic” complexity (following the pattern of genome:metagenome).

This involves looking at the number of distinct alleles/lings that exist across the entire population.

A simple trade pidgin used exclusively for simple transactions across disparate ethno-linguistic groups is very likely to have a low metalinguomic complexity. On the other hand, an ancient and entrenched language spoken by a large population is likely to have a high complexity (because they will have developed field-specific jargon, class indicator “alleles”, etc.).

Limits on Linguistic Complexity

Just as biological evolution places limits on the maximum information-theoretic complexity (measured in bits) of a population-wide metageneome, determined by factors such as

- Mutation rate
- Error-correction capacity
- Effectiveness of sexual selection as an error detection mechanism
- Ratio of reproducing to non-reproducing offspring and correlation between genetic fitness and reproductive success

and so on, there are also information-theoretic limits on the maximum metalinguomic complexity of a linguome population. The factors limiting maximum linguistic complexity of a language include

- The number of people speaking the language. If only a small population speaks the language, there will not be enough large subpopulations to develop mutually intelligible non-fixed lings (like field-specific jargon).
- The amount of linguistic capacity available to those individuals who speak the language. If the language is only used for narrow purposes (e.g. as a trade pidgin), people will not allocate large amounts of cognitive capacity to expanding it, so the complexity of the language cannot grow. Additionally, if the language is spoken primarily by a population of low general cognitive capacity, the language cannot grow to a high level of complexity because the population does not have “room” for extra non-fixed lings.

Fitness

In an evolutionary environment, we frequently speak of the (mal)adaptivity of genes/genomes and their fitness. We can do the same for lings/linguomes.

Monolingual traits

Some traits select on the level of individual lings.

Some factors that partially determine the relative fitness of a ling:

- Does it have an appropriate level of semantic density? Highly fit languages approach optimal codes, in that common phonemes/graphemes express

commonly referenced concepts. It would be maladaptive for an extremely common word like “the” to have an extremely cumbersome multi-syllabic representation, and conversely it would be maladaptive for an uncommon referent like “perigee” to have a compact monosyllabic representation like “gee”, stealing phonemic/graphemic bandwidth from more common referents like the letter “g” or the expression of surprise “gee”.

- Does it have a high information-theoretic efficiency? When categorizing the set of possible world-states, does it partition them in a semantically and computationally efficient way? “Light” and “Dark” are information-theoretically efficient; they partition the set of world-states (is an object light or dark?) into relatively even-sized sets (high negative log-likelihood) and is computationally efficient (the classification is easy to apply). On the other hand, a (fictional) word like “Blark”, referring to objects with an albedo precisely between 0.1 and 0.2, is maladaptive, because it neither efficiently partitions the set of world-states, nor is computationally efficient (easy to determine).
- Is it mutually communicable? Are there other linguomes that also contain this ling?
- Are the people who have this ling in their linguome people you *want* to communicate with? This is a surprisingly important consideration. Many of the French-derived semi-fixed lings in the English language (like “Beef”) exist almost entirely because they were inside the linguomes of high-status individuals in historical English-speaking populations. If you wanted to sound high-class, you had to adopt those lings. A more prosaic example of this is that when people in your field use a particular lingo, you had better be able to understand them, so you will absorb the lingo’s lings.

Polylinguic traits

Certain fitness characteristics, as in any biological organism, emerge from polygenic/polylinguic interactions.

For example, a common whole-linguome phenotypic characteristic of interest is the “difficulty” of a language. Some languages are understood to be unusually challenging to learn (conditioned on exposure to languages with similar ancestry), such as English or Japanese. Is being “difficult” an adaptive or maladaptive trait? It depends on the ecological niche.

For a trade pidgin, standard Japanese or English would be a poor choice, because it would be inordinately difficult to express simple trade-related concepts.

On the other hand, there are several significant second-order effects of linguomic complexity that can have unexpected adaptivity benefits. For example, an important social function of many languages is efficient and unforgeable social group identification. More difficult languages are more effective at this function, as there’s a more obvious gap between a speaker with a powerful grasp of the language (indicating some combination of intelligence or spare resources to

devote to language mastery) and a weak speaker (indicating, outside of mitigating circumstances like foreignness, some combination of lower intelligence or a lack of resources to allocate to language mastery). A simple language does not provide as clear a skill gradient, reducing its capacity as tool for social sorting.

Effects of Selective Pressure

Let's say we start with a high-capacity population (e.g. the British Isles) and let them develop a language. You end up with British English. Now, take the same language and airdrop it on a population with markedly lower capacity. What types of changes do we expect?

One of the most common changes is regularization of irregular inflections/conjugations. For example, in English, the copula "to be" is highly irregular and has many non-standard inflections. When linguistic capacity collapses, you see changes of the form

High-capacity: "I am going to the beach"

Low-capacity: "I be going to the beach"

We've replaced the irregular first-person inflection with the uninflected "dictionary" form.

Another common effect of linguistic complexity reduction is a transition from boolean semantic rules to scalar semantic rules.

Boolean semantics: "I did not do anything"

Scalar semantics: "I did not do nothing"

Under boolean semantics, the double-negation in the second sentence collapses and we're left with the shorter "I did something".

Under scalar semantics, rather than negation having a global effect on the entire proposition, each negative indicator simply adds to the level of "negativity" of the sentence. "I did not do nothing" becomes a stronger negation than "I did not do anything". The evaluation rules for boolean semantics are more complex than for scalar semantics, so it's no surprise that under capacity collapse, languages may switch from complicated-but-precise to simple-but-vague.