What Explains the Tower of Babel?

In this talk, I would like to arrive at a better explanation of why human language looks the way it does by drawing on the key insights from the language of biology. In particular, I would like to argue that understanding of human language and its variation – the "Tower of Babel" – can only be accounted for explicitly incorporating the *underlying* 'traits' of language into the Darwinian materialist view of individual variation and speciation, an epistemological stance that, to my mind, has not yet been fully appreciated or embraced.

So the idea I would like to talk about this: We must make the study of language *more* like biology, by *giving up* on two central idealizations that have held sway in modern linguistic theory over the past 50 years. Rather, we must attend to Darwin's notion that *variation* is of the essence, and the heart of speciation – and similarly different languages – is the conversion of the standing variation among individuals within an interbreeding group into variation between groups in space and time. Such a theory of evolution necessarily takes the variation between individuals as of the essence. Can we do the same for language?

Indeed, among the many puzzling questions about language, two are salient: First, why are there any languages at all, evidently unique to the human lineage, what evolutionary biologists call an "autapomorphy"? Second, why are there *so many* languages, apparently clumped into groups like biological species? These are in fact the basic questions of origin and variation that so occupied Darwin and other evolutionary thinkers and comprise modern biology's explanatory core: why do we observe *this* particular array of living forms in the world and not others? In this talk I will attempt to answer the second question, why are there so *many* languages, rather than just one: the Tower of Babel. A familiar and easy answer is that there are so many languages for the same reason there are so many different kinds of organisms – due to the potential for generating 'endless forms most beautiful' in the current evo-devo sense, as well as the drive to find distinguishable 'niches'. Indeed, it is this fact that is most often mentioned in the way that languages develop new vocabulary words to distinguish themselves from one another. As I am sure you are aware, philological analysis from the early 1800s to the present day, by uncovering rules governing phonological change, historical linguists reconstructed dead protolanguages such as Indo-European — an ancestral dialect spoken some 10,000 years ago that diverged into a wide variety of modern languages, including Hindi, Russian, Spanish, English and Gaelic. The crowning achievement of these early linguists was a family tree of languages that became an inspiration for Charles Darwin as he pondered biological evolution. [SLIDE]

However, this observation usually devolves into simply the *taxonomic observation* of distinct language families. It does not *explain* why there are certain groups and not others? *Why* do languages clump into groups? That is because, I believe, there is another, far, far *deeper* reason for the Tower of Babel. Not only do distinct languages wind up using different words with distinct sound patterns, e.g., *love* in English, *ama* in Portuguese; they differ along deeper dimensions, just a there is a complex mapping from genotype to phenotype that results in the morphological 'traits' that a taxonomist uses. To take an example literally closer to home, we know that our colleagues from Brasil would say *ele me viu* for "he saw me" rather than *ele viu-me*. (The first, Brasilian variant we will call "Classical Portuguese", and the second, "European, Modern Portuguese." But Why? This divergence is, in fact, part of what I will try to now *explain* as which I will now show is really almost like a genomic trait, rather than a word-structure one.

This important aspect of explanation – why we see certain languages and not others – *cannot* be captured if hold to the standard idealizations of modern linguistic theory. Consider for example a well-known statement of these idealizations, from Chomsky, 1965:

> "Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech-community, who knows and acquires its language perfectly…no cogent reason for modifying [this position] has been offered."

On the standard modern linguistic view then, not only is language communication perfect – no one ever misunderstands you, but also language learning is instantaneous and perfect. Impossible, of course, these idealizations are false, and linguists know they are false. They are meant to serve as Galilean idealizations: that one can *imagine* language understanding and learning *as if* it were perfect, and this does no harm to our construction of the 'true' theories of human language.

A moment's reflection ought to suggest that the idealization regarding perfect language learning leads to an immediate paradox, because it assumes that language is acquired from a single source (the 'homogeneous community' mentioned in the quote), and always works perfectly. But if children in the current generation always perfectly learned the language of their caretakers, and then, on growing up, as caretakers they in turn passed on their language to their own children, and so on, generation after generation, then languages would never change. Yet we know that they do. Here's an example. Why?

A common answer, true as far as it goes, is that by analogy with DNA replication, there must be some slight chance of *mis*learning the language of one's parents or peers. There is, however, a crucial *distinction* to be made with the Darwinian formulation in biology, and that is in the mode of inheritance -- the 'transmission' of information from one generation to the next. In the modern Darwinian synthesis this role is played by Mendelism. In the case of language, the transmission of information from parents and their surrounding social milieu to offspring is carried out by language acquisition. We do not need to say anything more at the moment about this than that for language acquisition the transition mechanism is simply that – any mechanism, i.e., any computable function whatsoever that accesses data from the environment (eg, sentences from parents or peers, suitably generalized, so we can think of this just as some datastream s_1, s_2, …) after some (finite) time, maps them onto some 'target' language (eg, Bangla, German, Japanese, French, ...).

Suppose then that we take the possibility of 'copying error' as in DNA – some probability of 'mislearning' a parent's language – to be source of divergence of languages over time. (This is the viewpoint expressed since the time of Darwin in the familiar analyses of how cognate words such as *water, wassen, vato* might 'diverge' over time, with recent work constructing phylogenetic trees of common ancestry, calculating divergence times, rates, and the like.) Suppose then that we set up the simplest model possible of such 'language transmission' along the following lines analogous to the biological situation, which are also deliberately aimed to be as general as possible (aside from one constraint that we shall come to in a moment):

- Assume there is some possible (finite or infinite) universe of possible languages; let us call this the *state space* of possible languages *L*. (So *L*={German, Dutch, Portuguese Bangla, Japanese, …}

- We further posit some (arbitrarily large) population of language speakers ('agents') at generation 1. Corresponding to the biological case, the agents might differ in the 'gene variants' (alleles), i.e., languages that they speak (or words they use for 'water'). We can assume some initial probability distribution of languages (traits) over agents. In particular, again following the simplest genomic situation, let's say that there is one 'gene' with 2 allele variations, language 1 (German say) and language 2 (Chinese). These are distributed in some frequency within the population, e.g., 80% of the agents might speak German, and 20% Chinese, etc. Call the language spoken by a particular agent in generation 1 l_1 (for the moment not distinguishing among languages).

- Each Agent transmits its language (trait) to a single language learner, via the learning algorithm *A*, by 'drawing' *n* example sentences in i.i.d (i independent draws) fashion given some probability distribution to create the linguistic experience *D* for the learner. *A* maps data (experience) to a selected language. Note that *A* can be *any* learning scheme: Bayesian estimation, some threshold

- The learner applies its learning algorithm *A* to the input data it receives, i.e., *A*(*D*), yielding the mature language it 'learns', or acquires, l_2. Then, this mature learner in turn becomes an agent to pass on l_2 in the next generation. To capture the notion that the transmission of information might not be entirely perfect (by analogy with the biological case), we define a *transition matrix* for any two languages *l* and *k* such that *T*[*l,k*], gives the probability that the learner exposed to the data from language *l* will learn language *k*, i.e.,
$$T[l,k] = prob[l \to k] = prob[A(D) = k \mid D \text{ generated by } P_l]$$

  Thus for example *T*[*l,l*] is the probability of correctly acquiring the parent agent's language, while if $l \neq k$ (over all other entries in the row), then the learner *incorrectly* acquires its parent's language. (The matrix *T* is stochastic: for a given row *l,* the column entries *k* must sum to 1.) So this is how we have modeled the notion of information mis-copying.

- This gives rise to an *iterated* trajectory of languages over generational timesteps: $l_1 \to l_2 \to l_3 \ldots$ Let us call this model *Iterated Learning* (IL).

- The IL model is very clearly now a *dynamical system*: we have described the state of a system at the next generation (time) *i*+1, a distribution of languages, as determined by an update equation given by *T*, the distribution of agents at the current generation *i*, the distribution of sentences each agent speaks, plus the update mapping *A*(*D*). The initial distribution of speakers of language *l* at generation *t*, $P^{(t)}(l)$, will evolve as follows:

We can picture this model as follows: [Slide]

Now let's see if this model, perhaps the simplest one imaginable that includes the possibility of slight *mis*learning, or information transmission/copying errors, as well as a simple notion of a non-homogeneous language community (two or more languages being spoken), actually *works* and makes sense. Does it do the job of incorporating Darwinian evolution properly, showing how languages might invariably 'clump' into distinct 'species' given what we know about language change and language stability? Perhaps surprisingly, the short answer is that it does not and cannot. It is completely dynamically insufficient, in that it can neither account for language stability, nor can it account for the empirically attested possibility of rapid language change, or the clustering of languages into groups. Let us see why, and see what part of Darwinism we have unfortunately missed, so we can modify the model and arrive at a better solution. By the end, I will hope to build enough of a model to explain part of the change from classical, "European" Portuguese to Brazilian Portuguese.

An example will illustrate the general result. Returning to our idealized world of just two languages (different alleles), say German and English assume some initial proportion of speakers of each language, say, 90% German speakers and 90% English speakers; call this initial percentage α_1 and $(1–\alpha_1)$. What do the entries of the corresponding 2 x 2 transition *T* matrix look like? The entry *T*(1,1) is the probability that a learner will receive examples from an agent speaking language 1, and correctly acquire language 1, e.g., *T*(1,1) gives the probability of learning German after being exposed to an German-speaking agent, call this *p.* *T*(1,2) is the probability that a learner will hear the agent speaking language 1, German but acquire language 2 instead (English), so *T*(1,2)= 1–*p* = 1–T(1,1) is the probability of mislearning German; call this probability epsilon (which we assume to be small). Thus German is effectively learnable with some small error. For simplicity, assume that this mislearning error probability is also the same for English learners exposed to the data from English speaking-agents, ε. Thus, *T*(2,2) and *T*(2,1) are the dual cases for the transition probability entries for learners successfully acquiring hearing English and mislearning German instead, *q* and ε, respectively.

Given this transition matrix *T,* when we iterate the update equation for the distribution of l_1 in the next generation, over and over again until we reach a fixed point (equilibrium condition) α_* where the percentage of l_1 speakers no longer changes, this yields the following equation:

$$\alpha_* = \frac{T(2,1)}{T(2,1)+(1-T(1,1))} = \frac{\varepsilon}{\varepsilon+\varepsilon} = \frac{1}{2}$$

I.e., a 50-50% mix of both languages! This holds *no matter* what the initial distribution of language speakers is. So, if there is even an infinitesimally small chance of mislearning either of the two languages, then from any initial mix, the system will converge to a 50-50 mix. Although each language is effectively learnable (with probability 1–ε), a homogeneous community *cannot* be maintained and degenerates to a mixture of language types over time. The only situation in which the single-agent, iterated learning model is stable is if learning is absolutely perfect, with zero error tolerance – not jettisoning the 'ideal speaker-hearer' idealization.

Although each learner is immersed in a potentially heterogeneous population, each learns only from a single individual, never reflecting the population variation. Different learners, of course, learn from potentially different individuals.

The case of Portuguese.

The main moral underscores the individual/population difference, and demonstrates the potential subtlelty of language change: we show that simply because an individual child will, with high probability, choose a particular grammar (European Portuguese) does not mean that all other grammars (e.g., Classical Portuguese) will be eliminated; rather, contrary to surface intuition, that is a property of the dynamical system and the population ensemble itself.

In this paper, we focus on a particular change in phonological and syntactic Portuguese recently discussed by Galves & Galves (1995). Roughly, over a period of 200 years, starting from 1800, ‘‘classical’’ Portuguese (CP) underwent a change in clitic placement. From the 16th century or before until the beginning of the 19th century, both proclitics and enclitics were possible in root declarative sentences (nonquantified subjects), as given by Galves & Galves (G&G) in examples (1) and (2), and in quantified subjects (3), which we will refer to henceforth via their reference numbers:

(1) Paulo a ama. (proclitic) Paulo her loves
‘Paulo loves her.’
(2) Paulo ama-a (enclitic) Paulo loves-her
‘Paulo loves her.’
(3) Quem a ama? (proclitic) Who her loves?
‘Who loves her?’

“During the 19th century a change affecting the syntax of clitic-placement occurred in the language spoken in Portugal.As a result, sentences like (1) became agrammatical and (2) remained as the only option for root affirmative sentences with non-quantified subjects. This change, however, did not concern sentences like (3) with quantified or *wh*-subjects in which proclisis was, and continues to be, the only option.” Galves & Galves, Tycho Brae project.

To each sentence we will assign (a) a morphological word sequence; (b) a stress contour; and (c) a syntactic structure. For example, again following G&G’s analysis, sentence type (2) will remain only in CP, while the two sentences (2)–(3) above will have different stress patterns for CP and EP. We omit a detailed description of the stress assignment and syntactic properties, as they are not necessary for our analysis. All we need to know is that G&G assume that the stress contours corresponding to sentence types (1), (2), and (3), which we denote simply as c_1 c_2 , c_3, follow a Markov chain description and, more importantly, govern the probability with which sentences are produced. (We are, of course, aware that this assumption of G&G may also be questioned; one might substitute any other more plausible relation between stress and sentence types — if any; this assumption is simply designed as a bridge to get the child from a presumably observable surface.) Thus, if two sentences have the same stress contour, then they will be produced with the same probability (given by the probability of the stress sequence according to

Markov production rules). In short, for the purpose of this paper, it is sufficient to assume that there are two and only two grammars (in accordance with Galves's assumptions): G_{CP}, denoting the grammar of Classical Portuguese (earlier) and G_{EP}, denoting the grammar of European Portuguese. Furthermore, the only data that is relevant (ignoring other aspects of the grammar) is as follows:

Classical Portuguese (CP-1) c_1: produced with probability p; (CP-2) c_2 produced with probability $1–2p$; and (CP-3) c_3 produced with probability p.

European Portuguese (EP-1) c_1: not produced; (2) (EP-2) c_2 produced with probability $1–q$; and (EP-3) c_3 produced with probability q.

Any (historically changing) population will now by assumption contain a mix of speakers of Classical and European Portuguese. The Classical Portuguese speakers produce the sentence types shown above with the probabilities (parameterized by p). The European Portuguese speakers produce the sentence types shown above with the probabilities (parameterized by q).

Thus we have defined (1) the class of grammars (1) $G=\{G_{EP}, G_{CP}\}$; (2) Probabilities with which speakers of G_{EP} and G_{CP} produce sentences (parameterized by p and q). We therefore can derive the evolutionary consequences on the population for a variety of learning algorithms. We first consider a probabilistic, Maximum Likelihood Method: to choose between CP and EP given some input sentences (conditioned on stress patterns), pick the language (grammar) that maximizes the probability of generating the available data (surface forms). This is probably the simplest probabilistic learning algorithm and leads to the following:

Learning Algorithm 1
1. Draw n examples (sentences).
2. Compute likelihoods, i.e., $P(S_n \mid G_{CP})$ and $P(S_n|G)$
3. Use the Maximum Likelihood Method to choose between the two grammars.

As discussed earlier, to calculate the historical dynamics we must be able to analyze the behavior of the learning algorithm, i.e., characterize eq. 1. For the analysis of the algorithm, we assume that sentences are drawn in i.i.d. fashion according to a distribution dictated by their stress contours as indicated in the earlier section.

First, consider the form of the likelihoods. Let the example sentences be $S_n=\{s_1, s_2, \ldots, s_n\}$. Due to the i.i.d. assumption, $P(S_n \mid G_{CP})$ is given by $\Pi_{i=1}$ to n $P(s_n \mid G_{CP})$. Suppose that the sample consists of a draws of c_1, b draws of c_3, and $n–a–b$ draws of c_2. Then the following is clear:

$$P(S_n \mid G_{CP}) = p^a (1-2p)^{n-a-b} p^b$$
$$P(S_n \mid G_{EP}) = (0)^a (1-q)^{n-a-b} q^b$$

Consequently, the individual child, following the Maximum Likelihood Method will choose the grammar *EP* (G_{EP}) **only if** (1) no instances of c_1 occur in its sample; and (2) c_2 and c_3 occur in numbers so that $q^b \ (1–q)^{n\_b} > (1–2p)^{(n–b)} p^b$. There are 3 cases to consider.

Case 1. $p < q < 2p$. *Decision Rule:* For this case, it is possible to show that the child (following the Maximum Likelihood Rule) always chooses *GEP* if no instances of c_1 occur. This is simply because $1–q > 1–2p$; and $q>p$.

Equation 1 and Population Update: Suppose that the proportion of speakers of *GCP* in the *i*th generation is α_i Then the probability of drawing c_1 is given by $\alpha_i p$. Consequently, the probability of drawing a set of *n* examples without a single draw of c_1 is $(1 - \alpha_1\, p)^n$. This is of course the probability with which the individual child chooses the grammar of European Portuguese, G_{EP} Thus the update rule has the following form:

$$\alpha_{i+1} = 1-(1-\alpha_i p)^n$$

Case 2. *$q < p < 2p$ Decision Rule:* In this case, the Maximum Likelihood Rule reduces to the following. Choose G_{EP} *if and only if* (1) $a = 0$, i.e., no instances of c_1 occur; and (2) $b < n\Upsilon$ where $\Upsilon = \log(1\text{-}2p)/\log(1\text{-}q)/(1\text{-}2p) + \log(p/q)$. For all other data sets, choose G_{CP}.

Equation 1 and Population Update: As usual, let there be α_i proportion of the previous generation speaking G_{CP}. It can be shown that events (1) and (2) above occur with probability $\Sigma_{k=0 \text{ to } n\Upsilon}$ *n choose k,* $P^k Q^{(n-k)}$ where $P = \alpha_i p+(1-\alpha_i)q$, as shown below.

$$\alpha_{i+1} = 1-\sum_{k=0}^{n\gamma}\binom{n}{k} p^k Q^{(n-k)},\ Q = \alpha_i(1-2p)+(1-\alpha_i)(1-q)$$

$$P = \alpha_i p+(1-\alpha_i)q$$

Case 3. *$p < 2\,p < q$. Decision Rule:* The MLE reduces to: choose G_{EP} iff (1) a=0; and (2) $b > n\Upsilon$. Otherwise, choose G_{CP}. The update rule has the following form:

$$\alpha_{i+1} = 1-\sum_{k=n\gamma}^{0}\binom{n}{k} p^k Q(n-k),\ Q = \alpha_i(1-2p)+(1-\alpha_i)(1-q)$$

$$P = \alpha_i p+(1-\alpha_i)q$$

System Evolution

We have shown above how the behavior of the population can be characterized as a dynamical system and have derived the update rules for such a system for a Maximum Likelihood learning algorithm The dynamical system captures the evolutionary consequences of this particular learning algorithm. In this section we describe its evolutionary properties, and see how they mesh with observed cognitive (historical) trends.

Three Cases

Case 1. $p < q < 2p$

1. $\alpha=0$ is a fixed point, i.e., if the initial population consists entirely of European Portuguese speakers, it will always remain that way. Furthermore, if $np<1$, then this is a *stable* fixed point. It is also the only fixed point between 0 and 1. Thus in this case a population speaking entirely Classical Portuguese would gradually be converted to one speaking entirely European Portuguese.
2. If $np > 1$, then $\alpha=0$ remains a fixed point but now becomes unstable . For this case, an additional fixed point (stable) is now created between 0 and 1. All initial

population compositions will tend to this particular mix of G_{CP} and G_{EP} speakers. Figure 1 shows the fixed (equilibrium) point as a function of *n* and *p*.

Case 2. $q < p < 2p$

1. Unlike case 1, the dynamical evolution now depends upon both *p* and *q* in addition to *n*.
2. It is easily seen that α=0 is no longer a fixed (equilibrium) point (unless p=q). Consequently, populations, irrespective of their initial composition, will *always* contain some speakers of Classical Portuguese.
3. It is possible to show that there is exactly one fixed (*stable*) point and all initial populations will tend to this value. Shown in fig. 2 is plot of the fixed point as a function of *q* and *p* for a fixed value of *n*. Notice the multiple ridges in the profile suggesting sensitivity to the value of *q* around some critical points. Shown in fig. 3 is a plot of the fixed point as a function of *p* for various choices of *n* keeping *q* fixed at 0.1.

Case 3. $q < p < 2p$. Like case 2, the dynamical evolution depends upon both *p* and *q* in addition to *n*. 2. Again, it is easily seen that α=0 is no longer a fixed point. Therefore, the speakers of Classical Portuguese can never be eliminated altogether for *p* and *q* in this range.

What are the important conclusions from this analysis? In short, children using the Maximum Likelihood Rule will choose *GEP* over *GCP*. However, a dynamical systems analysis must be carried out to see if that will suffice to ''wipe out'' Classical Portuguese. Only in case 1 will Classical Portuguese be lost completely (provided *p* ` 1a*n*). *In all other cases* , there will always remain some speakers of Classical Portuguese within the community. In fact, the evolutionary properties can be quite subtle.4 Consider the following three example cases.

Example 1 *Let p =0.05, q*=0.02, *and n*= 4. *In this case, if the parental generation were all speaking Classical Portuguese then a simple computation shows that the probability with which the child would pick* G_{EP} (α=1) European *Portuguese, is* 0.66. *i.e., it is greater than one-half. Thus,* **in spite of the fact** *that the* **majority** *of children choose the grammar of European Portuguese, the speakers of Classical Portuguese will never die out completely. In fact, the fixed point is 0.11. Roughly 11 percent of the population will continue to speak Classical Portuguese.*

Example 2 *Let p* =0.05, *q*= 0.06 *and n* =8. *If this were the case, and the parental generation all spoke Classical Portuguese, it turns out that the probability with which the individual child would pick GEP would again be* 0.66 *However, now the speakers of classical Portuguese would all be lost and the population would move to its stable, fixed point containing only speakers of European Portuguese.*

Example 3 *If p* =0.05, *q*= 0.06 *and n*=21 *however, it is easily seen that Classical Portuguese speakers can never be completely lost.*

Since a batch algorithm is presumably psychological unreal (due to memory limitations), one could substitute, as we have done, a memoryless algorithms, such as local gradient ascent (Gibson & Wexler's ''Triggering Learning Algorithm'' or TLA, 1994). Due to reasons of space, we leave a detailed presentation of the results of this modification to one side, and simply note that one obtains yet a different historical dynamic. It is possible to prove that in this case, the population evolves according to the update rule:

Here $$\alpha_{i+1} = 1 - \frac{1}{2}(1 - \alpha_i p)^n$$

Here alpha_i and alpha_i+1 are the proportion of the population speaking Classical Portuguese in generation *i* and *i*+1 respectively. As usual, *n* is the number of examples drawn. Here, note that CP speakers can *never* be eliminated to less than 1/2 of the population. Consequently, one is able to see immediately that the TLA does *not* have the right evolutionary properties to explain the change from Classical to European Portuguese. Second, it is possible to show that there is exactly one (stable) fixed point (between 1a2 and 1) to which such a system evolves, for various values of *n* and *p*.

What is going wrong?
What is the nature of this trajectory – that is, of this dynamical system?
Linear
Markov transition matrix.
Implications.
This has the non-homogeneous part.
What does it lack???
It lacks learning from *multiple* individuals. Social learning.
It lacks the 'non-instantaneous' part
Maturation time is crucial.

Symmetry breaking creates language species. Ising model: no phase transition in 1D case;

Discuss the SL model, show the graph. Stable Attractor basins. Result is to form two 'species' of language: one German-like (Scandanavian), one English-like.
Symmetry-breaking in the graph: jumps in attractor basin. Similarity to Eisen model

New criterion for linguistic theory, drawn from biology: must also be able to describe the correct dynamical trajectories over time. And the correct divergences of languages into the observed 'species.'

What is a parameter? A
What are the lessons?

We note that this shift from the viewpoint of the {\it individual}
acquisition of a grammar from data generated from a {\it single}
source to the population view is quite parallel to the fundamental
epistemological shift that took place with Darwin's introduction of
population level reasoning in biology at the time of {\it Origin of

Species\/}. Darwin insisted that in his theory individual variation was of the essence: for Darwin, biological evolution operates on two variational levels -- variation among the parental generation, and variation in the offspring generation. This is no mere philosophical quibble without theoretical or empirical consequences. Rather, it is central. We argue in this paper that the proper formulation of the full population viewpoint requires, as Lewontin (1974)\nocite{Lewontin74} notes, that one incorporate variation into both parental and offspring generations.\footnote{``$\ldots$ where is the revolution that Darwin made? The answer is that the essential nature of the Darwinian revolution was neither the introduction of evolutionism as a world view (since historically that is not the case) nor the emphasis on natural selection as the main motive force in evolution (since empirically that may not be the case), but rather the replacement of a metaphysical view of variation among organisms by a materialistic view (Lewontin, 1973). For Darwin, evolution was the conversion of the standing variation among individuals within an interbreeding group into variation between groups in space and time. Such a theory of evolution necessarily takes the variation between individuals as of the essence. Ernst Mayr has many times pointed out, especially in Animal Species and Evolution (1963), that this emphasis on individual variation as the central reality of the living world is the mark of modern evolutionary thought and distinguishes it from the typological doctrine of previous times."
1974, pp.3--5.}

Symmetry breaking in language: Curie quote
On web.
What is a parameter? Extracting parameters from 'control theory' points (stickleback)
VO-OV parameter head-first, head-final (selection)
Read about parameters.
Graphs.
Empiricism and internal state (rationalism)'
Conclusion

In the domain of language acquisition, variation, and change, this amounts to admitting both multiple possible languages – this corresponds to variation at the 'parental' level – and multiple learners each of whom acquires a potentially different grammar – the learners and the language they develop corresponding to the 'offspring' generation.