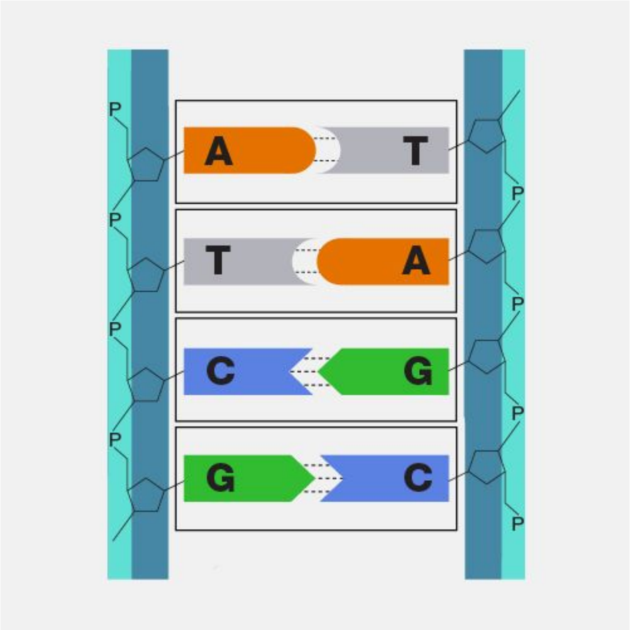
Computational and Evolutionary Aspects of Language

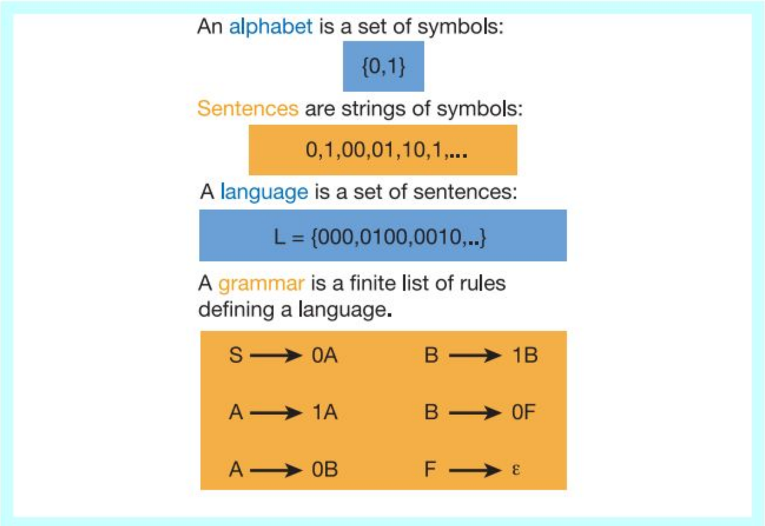
Nowak, Komovara, & Niyogi, 2002

Presented by Logan Swanson
Nov 15 2022

Generative Systems



DNA → Genetic Evolution

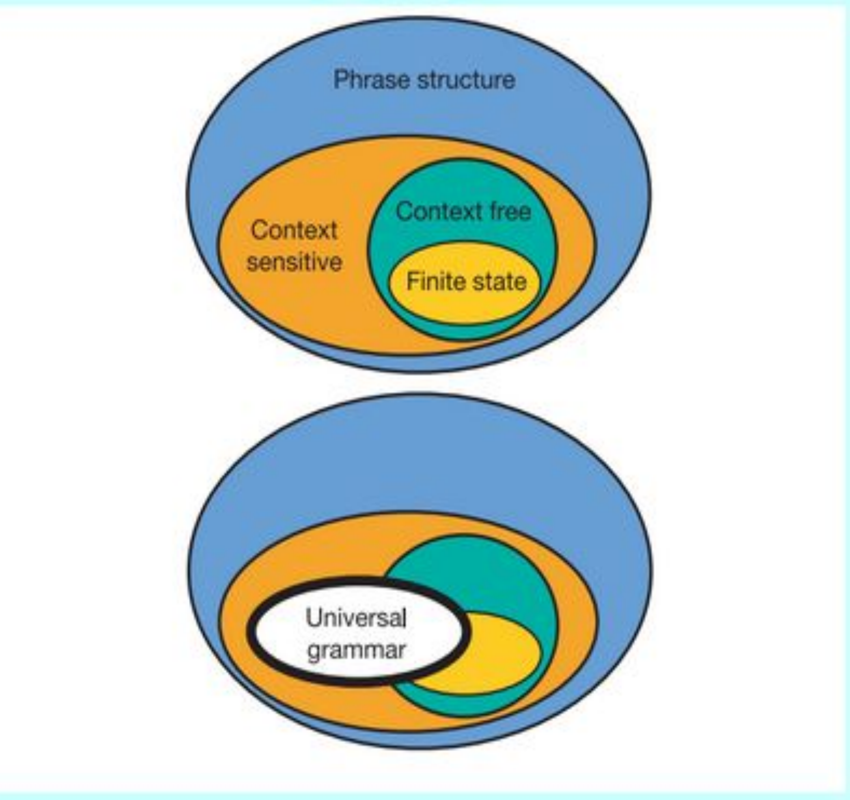


Language → Cultural Evolution

“The difference between ‘learning’ and ‘memorization’ is the ability to generalize beyond one’s own experience to novel circumstances.”

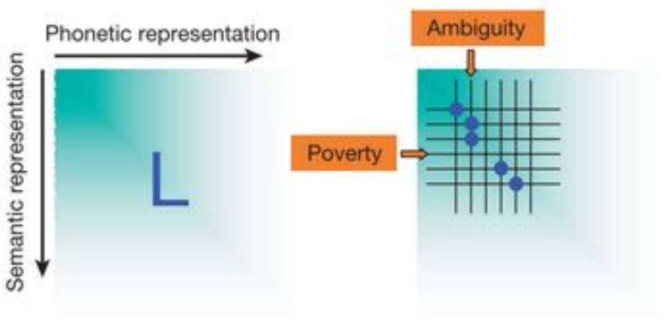
Chomsky Hierarchy

Where do natural languages fall?



Approaches to the Learning Algorithm

Sound / Meaning Pairs



(Lexicon)

Principles & Parameters

[+ V2]
[- Spec Initial]
[- v to T raising]
…

(Syntax)

Optimality Theory

1. *#CC
2. *NC#
3. …

(Phonology)

Evolution of Language

- Which genetic modifications led to changes in brain structures that were decisive for human language?
- How do we model the population dynamics of language change?
- How do we model change to UG itself?

Evolution of Language with Constant UG

Assumptions:

- All individuals have the same UG, and this does not change over generations
- UG is a finite concept class
- Each individual in the population speaks some language $L_i \in UG$
- Fitness of a language determines how likely children are to hear it as input

Language Fitness

“The fitness of a given language is equal to the weighted average of the communicative payoff between that language and all languages in the community, including itself”

$$f_i = \sum_{j=1}^{n} x_j F_{ij}$$

Fitness of language L_i

Fraction of the population speaking L_j

Communicative payoff of speaking L_i to a listener using L_j

Example

	L_1	L_2	L_3
L_1	1	0.2	0.2
L_2	0.2	1	0.2
L_3	0.2	0.2	1

Population: $x_1 = 0.7$, $x_2 = 0.2$, $x_3 = 0.1$

$f_1 = 0.7(1) + 0.2(0.2) + 0.1(0.2) = 0.76$

$f_2 = 0.7(0.2) + 0.2(1) + 0.1(0.2) = 0.35$

$f_3 = 0.7(0.1) + 0.2(0.2) + 0.1(1) = 0.28$

Language Dynamical Equation

"The change in the proportion of a population speaking L_j over time depends on how likely a child is to learn L_j given input from each language present in the population, adjusted for the frequencies of each language and their fitness within the current distribution"

Fraction of population speaking specific language

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i(\mathbf{x}) Q_{ij} x_i - \phi(\mathbf{x}) x_j$$

Change in proportion of population speaking L_j over time

Fitness of L_i

Probability that a child will develop L_j based on input from L_i

Average fitness of population

Example: Cont.

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i(\mathbf{x}) Q_{ij} x_i - \phi(\mathbf{x}) x_j$$

Population: $x_1 = 0.7$, $x_2 = 0.2$, $x_3 = 0.1$

Recall: $F_{ij} = 1$ for $i=j$, 0.2 otherwise

$f_1 = 0.76$

$f_2 = 0.35$

$f_3 = 0.28$

$\phi(\mathbf{x}) = 0.7(0.76) + 0.2(0.35) + 0.1(0.28)$
$= 0.63$

Assume $Q_{ij} = 1$ for $i=j$, 0 otherwise

$\Delta x_1 = 0.76(1)(0.7) + 0 + 0 - (0.63)(0.7) = 0.091$

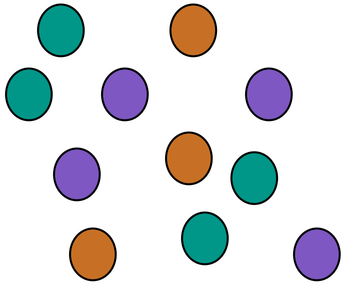
$\Delta x_2 = 0 + 0.35(1)(0.2) + 0 - (0.63)(0.2) = -0.056$

$\Delta x_3 = 0 + 0 + 0.28(1)(0.1) - (0.63)(0.1) = -0.035$

New Population: $x_1 = 0.791$, $x_2 = 0.144$, $x_3 = 0.065$

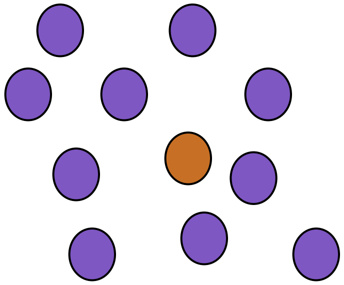
Equilibrium Solutions

Randomness



- All languages occur with similar frequencies
- Always a possible equilibrium
- Only stable when learning has a high error rate (ie Q is far from the identity matrix)

Coherence



- One language dominates the population
- Only possible & stable when error is below a certain threshold (Q is close to identity matrix)

Memoryless Learner

1. Select a random hypothesis from the concept class
2. If you hear a sentence that is not compatible with your hypothesis, randomly select another hypothesis from the class
3. After N sentences, permanently fix whatever hypothesis you are currently holding

In the special case where $F_{ij} = a$ when $i \neq j$, the following must be true in order to achieve linguistic coherence: $N > C_1|UG|$, where C_1 is some constant depending on a.

Batch Learner

1. Memorize N sentences
2. Choose the language most compatible with all N sentences. This is your permanent grammar.

In the special case where $F_{ij} = a$ when $i \neq j$, the following must be true in order to achieve linguistic coherence: $N > C_2 \log(|UG|)$, where C_2 is some constant depending on a and $|UG|$.

The claim made here is that this together with the memoryless algorithm results place bounds on the size of UG.

Evolution of UG

- Occurs on a much longer timescale than language evolution
- Impacted by selective pressure for UGs that can induce coherence

UG dynamical equation

"The change in prevalence of some language j in concept class U_J over time is equal to the likelihood that other UGs mutate into U_J multiplied by the rate at which offspring learn language j from inputs from languages in those UGs"

$$\frac{dx_{Jj}}{dt} = \sum_{I=1}^{M} W_{IJ} \sum_{i=1}^{n} f_{Ii} Q_{ij}^{J} x_{Ii} - \phi x_{Jj}$$

- Not much currently known about behavior of this system
- In the limit with no mutation, we find that equilibria converge to a single UG
- When this is relaxed, sometimes multiple UGs can be observed

Key Takeaways

- Many language changes are selectively neutral. This allows for linguistic variation and ongoing change.
- Deeply understanding the behavior of algorithms and the concept classes they belong to is crucial for mathematical insight about language evolution
- We can use the idea of linguistic coherence to bound the types of concept classes and algorithms that we are willing to consider