The Computational Nature of Language Learning and Evolution

Partha Niyogi

The University of Chicago

A Linguistic Fact of *English*

1 (a) He ran from there with his money.

1 (b) He his money with there from ran. $(*)$

A Linguistic Fact of *English*

1 (a) He ran from there with his money.

1 (b) He his money with there from ran. $(*)$

Linguistic Experience $\mapsto$ Linguistic Knowledge

The Logical Problem of Language Acquisitio

$\mathcal{G}$ $g_t \in \mathcal{G}$ target grammar

S_n $+$ve examples

$\mathcal{A}$ Learner (Child)

$$\mathcal{A}(S_n) = h_n$$

The Logical Problem of Language Acquisitio

$\mathcal{G}$ $g_t \in \mathcal{G}$ target grammar

S_n $+$ve examples

$\mathcal{A}$ Learner (Child)

$$\mathcal{A}(S_n) = h_n$$

Learnability $h_n \rightarrow g_t$

Gold (1967); Valiant (1984)

Principles and Parameters

Swahili

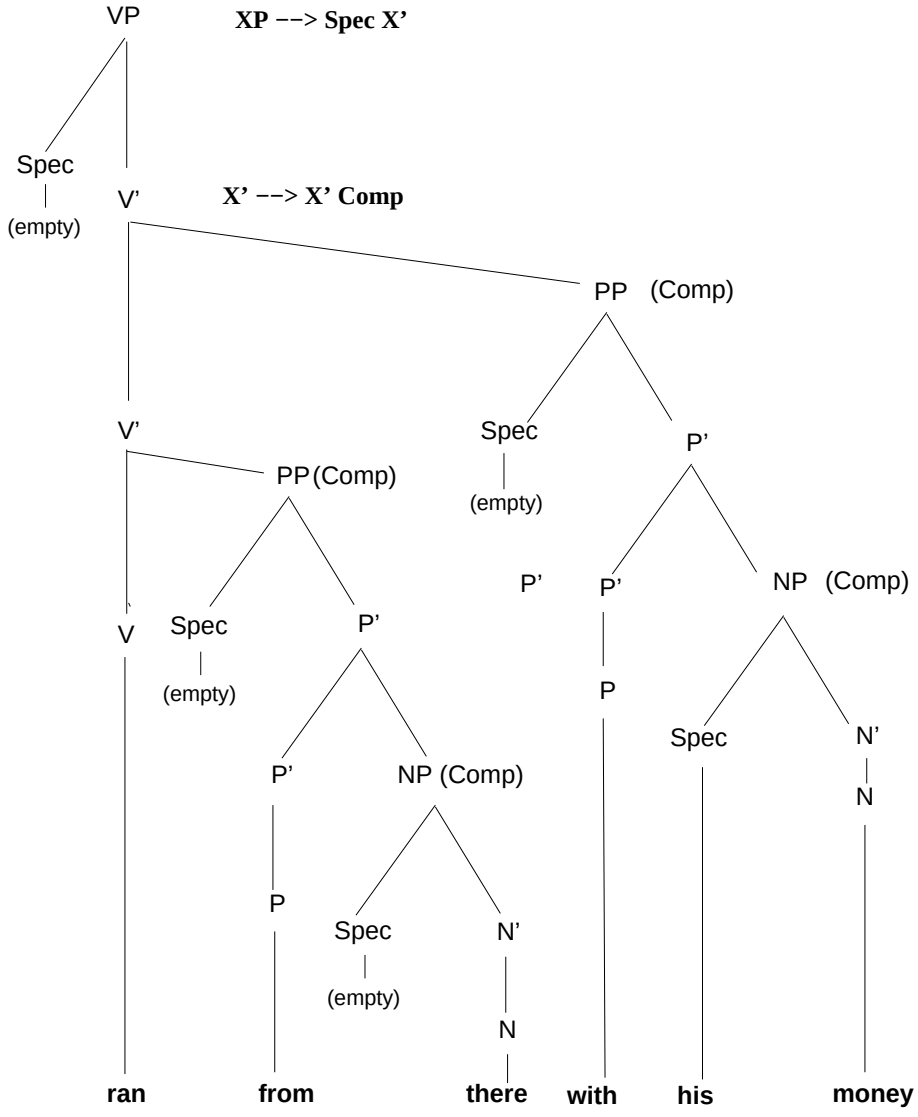
Chinese Hindi

$$\mathcal{G}(p_1, \ldots, p_n)$$

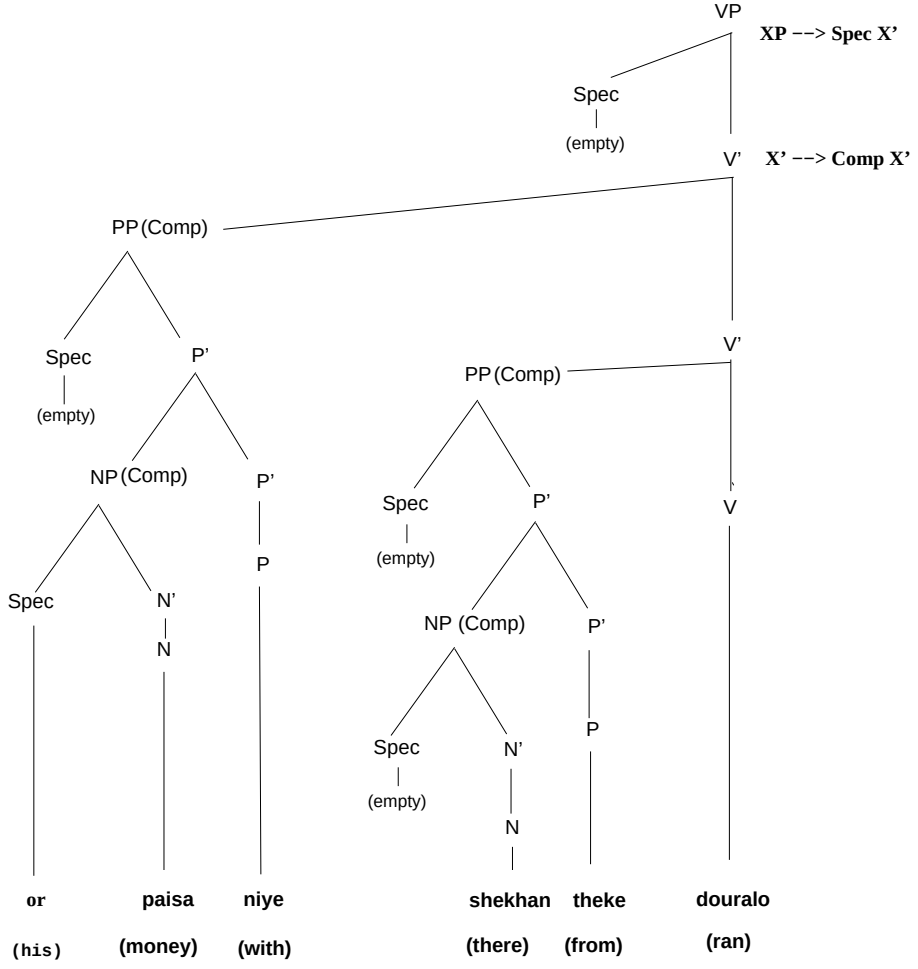
English French

Warlpiri

X-bar Theory: English



X-bar Theory: Bengali



The (Il)-Logical Problem of Language Chan

Langagis, whos reulis ben not writen, as ben Englisch, Frensch and many otheres, ben channgid withynne yeeris and countrees that oon man of the oon cuntre, and of the oon tyme, myghte not, or schulde not kunne undirstonde a man of the othere kuntre, and of the othere tyme; and al for this, that the seid langagis ben not stabili and fondamentali writen

Pecock (1454) *Book of Feith* (from Roberts, 1993)

The Evolution of English

Her ... Aelfred cyning ... gefeaht wid ealne, here, and hine

Here Alfred king fought against whole army and it

geflymde and him aefter rad od pet geweorc, and paer saet

put to flight and it after rode to the fortress and there camped

XIII niht, and pa sealde se here him gislas and myccle

fourteen nights and then gave the army him hostages and great

adas, pet he of his rice woldon, and him eac geheton

oaths that they from his kingdom would [go] and him also promised

pet heora cyng fulwihte onfon wolde, and hi paet gelaston

that their king baptism receive would and they that did

More Old English

pa Darius geseah paet he oferwunnen beon wolde

then Darius saw that [he conquered be would]

(Orosius 128.5)

& him aefterfylgende waes

and [him following was]

(Orosius 236.29)

Nu ic wille eac paes maran Alexandres gemunende beon

now I will also [the great Alexander considering be]

(Orosius 110.10)

Complex Constructions

ondraedende paet Laecedemonie ofer hie ricsian mehten swa hie aer dydon

dreading that Laecedemonians over them rule might as they before did

“dreading that the Laecedemonians might rule over them as they had done in the past”

(Orosius 98.17)

peh ne geortriewe ic na Gode paet he us ne maege gescildan

although not shall-distrust I never to-God, that he us not can shield

“although I shall never distrust God so much as to think he cannot shield us”

(Orosius 86.3)

Ogden Nash, 1962

Farewell, farewell to my beloved language
Once English, now a vile orangutanguage

Evolution and Learning

...if languages were learnt perfectly by the children of each generation, then language would not change: English children would still speak a language as old atleast as Anglo Saxon and there would be no such languages as French or Italian.

(H. Sweet, 1899)

Language Change

ISSUES/QUESTIONS

1. Group Level Description: *How does one model populations of linguistic agents?*
2. Time Course: *How fast do languages change? Can one predict their possible evolutionary patterns?*
3. Directionality: *When two language types come together, in which direction will the children evolve?*

HISTORICAL PHENOMENA

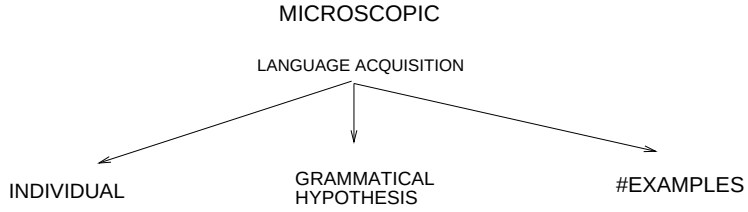
1. **Syntax:** Change in word order in English, French, Portuguese, etc.
2. **Phonology:** (a) Change in metrical stress from Proto-IndoEuropean to modern Greek (b) The Great Vowel Shift in English...
3. **Creoles:** Rapid language formation. Do all creoles have similar properties?
4. **Language Typology:** What are language types? How are they distributed? How do they change?

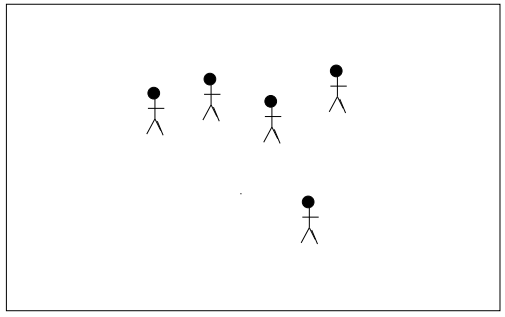
Language Evolution

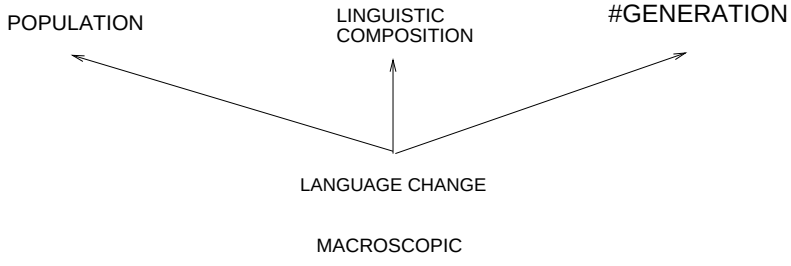
1. Origin of Language: *How did combinatorial, recursive structures emerge?*
2. Communicative Efficiency: *What is the role of communicative efficiency and natural selection?*
3. Communicative Coherence: *How do shared communication systems arise by self organization?*
4. Diversity: *How did the diversity of natural communicative systems evolve?*

Birds, Bees
Whales, Dolphins
Primates, Humans

Population Linguistics







Timeline of Inquiry

18th century	William Jones	
	(Indo-European Thesis)	
19th century		Charles Darwin
20th century	Linguistic Structure	Genetic Code
	(Generative Grammar)	DNA (Molecular)

Timeline of Inquiry

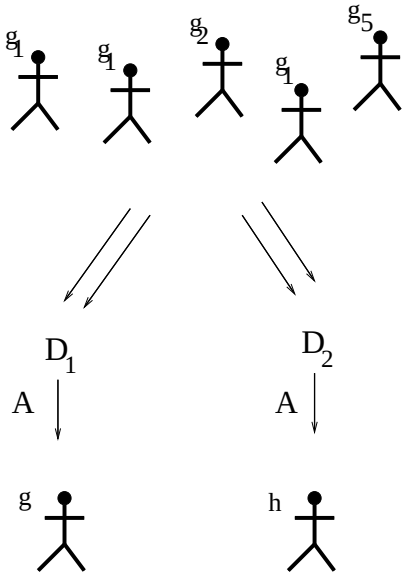
18th century	William Jones	
	(Indo-European Thesis)	
19th century		Charles Darwin
20th century	Linguistic Structure	Genetic Code
	(Generative Grammar)	DNA (Molecular)

The formation of different languages and distinct species, ... are curiously parallel.

(Charles Darwin, *Descent of Man*, 1871)

Evolution in Linguistics and Biology

Language Evolution **and** *Biological Evolution*



grammatical variation in adults

genetic variation in adults

transmission via learning

transmission via inheritance

grammatical variation in children

genetic variation in children

Natural Selection ??

Major Insights

1. Different learning algorithms have different evolutionary consequences.

   evolutionary criteria in addition to learnability

   learning in heterogeneous populations

2. Phase transition phenomena in linguistic evolution.

   subtle changes in frequency may lead to dramatic changes in language

3. Natural selection, Social connectivity, and the Emergence of Language.

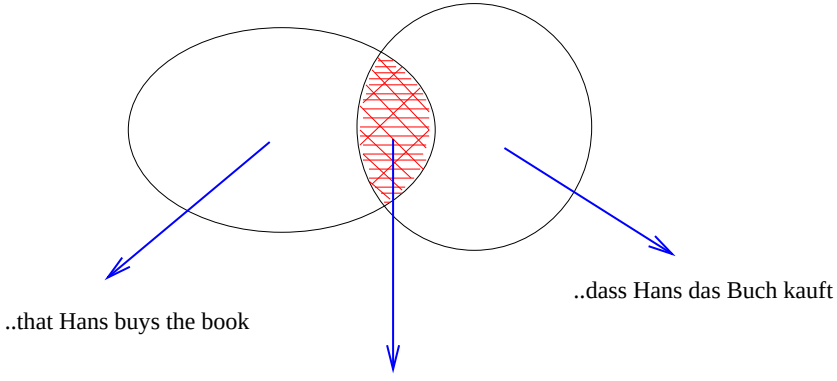
   conditions for a shared language to emerge

The Basic Framework

1. $\mathcal{L} = \{L_1, \ldots, L_n\}$
2. $(x_1, x_2, \ldots, x_n) \in \Delta^{n-1}$
3. $\{P_1, P_2, \ldots, P_n\}$
4. $\mathcal{A} : Data \rightarrow \mathcal{L}$
5. k: maturation time
6. $F : \Delta^{n-1} \rightarrow \Delta^{n-1}$

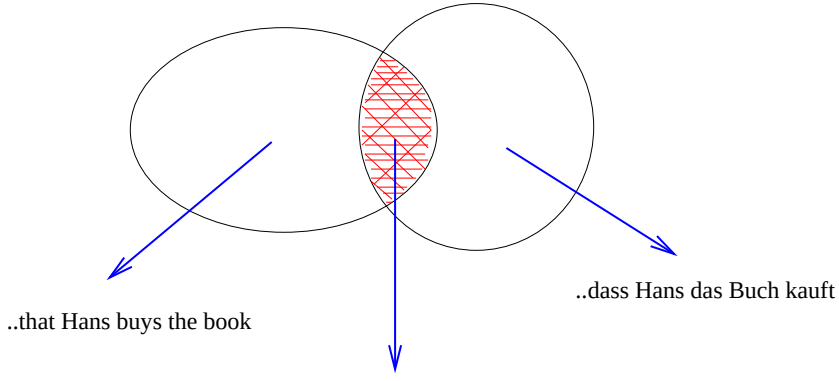
Two Language Models

$$\mathcal{L} = \{L_1, L_2\}$$



Two Language Models

$$\mathcal{L} = \{L_1, L_2\}$$

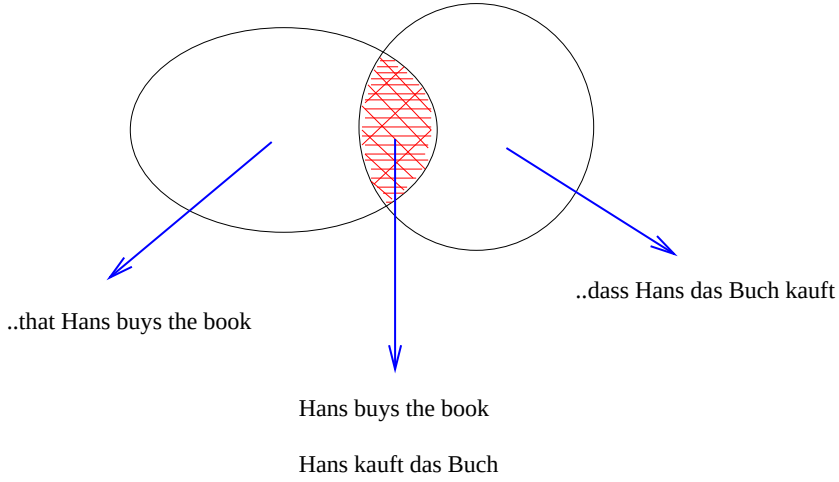


$$\mathcal{P} = \{P_1, P_2\}$$

where $P_1[L_1 \cap L_2] = a$ and $P_2[L_1 \cap L_2] = b$

Two Language Models

$$\mathcal{L} = \{L_1, L_2\}$$



$$\mathcal{P} = \{P_1, P_2\}$$

where $P_1[L_1 \cap L_2] = a$ and $P_2[L_1 \cap L_2] = b$

$$\mathcal{A} = ??$$

Triggering Models

1. Start with arbitrary hypothesis.
2. Receive new example sentence, s_n
3. *if* (s_n parsed)
   then go to 2
   else change hypothesis

The Logic of Language Evolution

Consider a typical child

$$s \sim x_t P_1 + (1 - x_t) P_2$$

After k examples,

$$\mathbb{P}[\mathcal{A}(D) = L_1] = f(x_t, a, b)$$

Therefore in the next generation,

$$x_{t+1} = f(x_t, a, b)$$

Population Dynamics

$$x_{t+1} = \frac{B + \frac{1}{2}(A - B)(1 - A - B)^N}{A + B}$$

$$A = (1 - b)(1 - x_t) \qquad B = (1 - a)x_t$$

Population Dynamics

$$x_{t+1} = \frac{B + \frac{1}{2}(A - B)(1 - A - B)^N}{A + B}$$

$$A = (1 - b)(1 - x_t) \qquad B = (1 - a)x_t$$

$$x_{t+1} = \frac{x_t(1 - a)}{(1 - b) + x_t(b - a)}$$

$a < b$	$a = b$	$a > b$
$x_t \to 1$	no change	$x_t \to 0$

Appreciating the Bifurcation

Let $x = 0.99$ and $a = 0.11$, $b = 0.10$.

$$s \sim xP_1 + (1 - x)P_2$$

Appreciating the Bifurcation

Let $x = 0.99$ and $a = 0.11$, $b = 0.10$.

$$s \sim xP_1 + (1-x)P_2$$

In the experience of the typical child:

1. Overwhelmingly many *triggers* for L_1
88% unique L_1 parse
$< 1\%$ unique L_2 parse

2. With high probability 0.9898 the child acquires L_1

Yet....

Cue Based Models

$$C \subset L_1 \setminus L_2$$

$$p = P_1(C)$$

1. Receive N example sentences.
2. Let k be $\#$ of cue sentences.
3. *if*

$$\frac{k}{N} > \tau$$

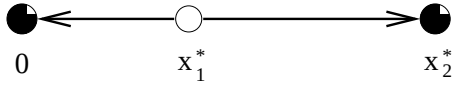
choose L_1.

Population Dynamics

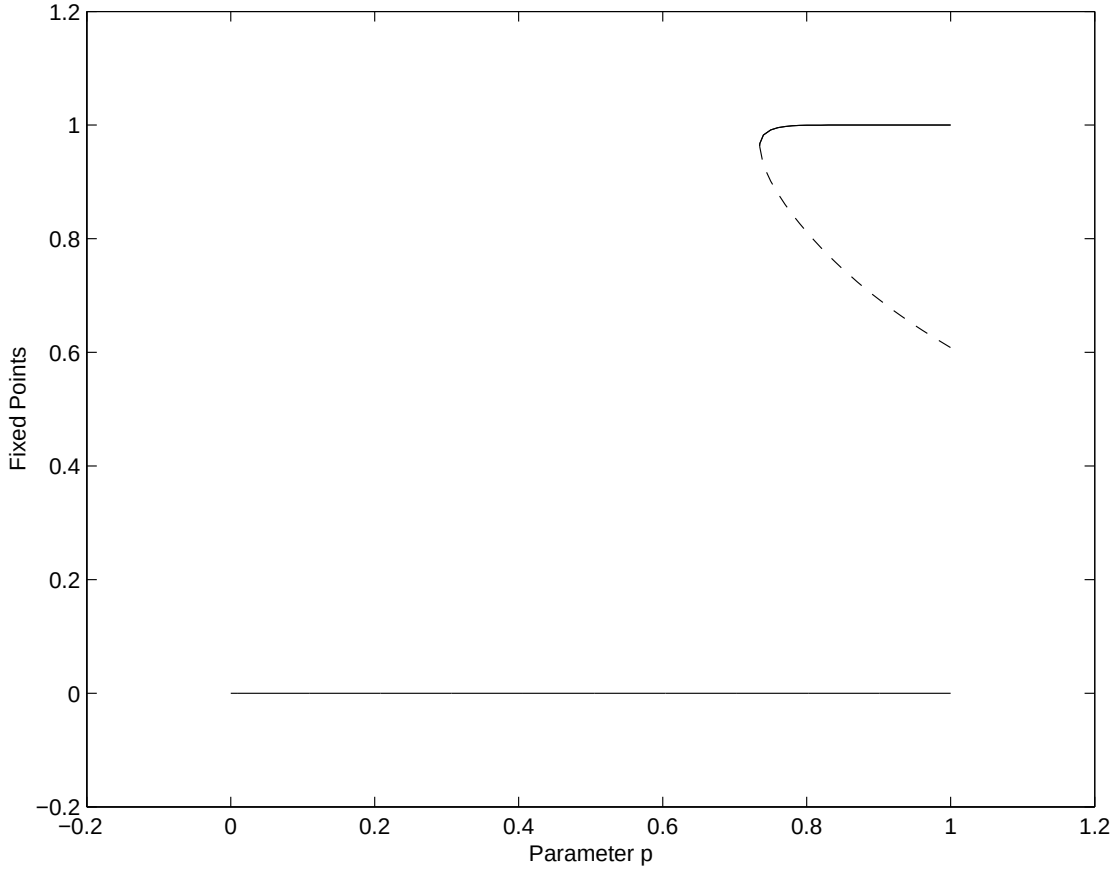
$$x_{t+1} = \sum_{i=N\tau}^{N} (px_t)^i (1 - px_t)^{N-i}$$

$x = 0$ (stable) $x = x_1^*$ (unstable) $x = x_2^* > x_1^*$ (stable)

$x = 1$ not equilibrium



The Bifurcation



Middle English

NORTHERN GRAMMAR	SOUTHERN GRAMMAR
(Scandinavian)	(Saxon)
$+V2$	$-V2$
SV, SVO	SV, SVO
OVS	S V O1 O2
S V O1 O2	Adv S V
O1 V S O2	Adv S V O1 O2
O2 V S O1	Adv S V O1 O2
Adv V S O	S Aux V
etc.	etc.

Evolution and Communicative Fitness

In the evolution of languages the dicarding of old flexions goes hand in hand with the development of simpler and more regular expedients that are rather less liable than the old ones to produce misunderstanding.

Otto Jespersen (1922)

The Emergence of Coherence – Fitness

1. $\mathcal{L} = \{L_1, L_2, \ldots, L_n\}$

2. $F(L_i, L_j)$ – mutual intelligibility

$F(L_i, L_i) = 1$ and $F(L_i, L_j) = a$

3. Q_{ij} – parent $L_i \rightarrow$ child L_j

$Q_{ii} = q$ and $Q_{ij} = \frac{1-q}{n-1}$

Population Dynamics

State $\{x_1, \ldots, x_n\} \in \Delta^{n-1}$

Population Dynamics

State $\{x_1, \ldots, x_n\} \in \Delta^{n-1}$

Fitness $f_i = f_0 + \sum_{j=1}^{n} F(L_i, L_j) x_j$

Population Dynamics

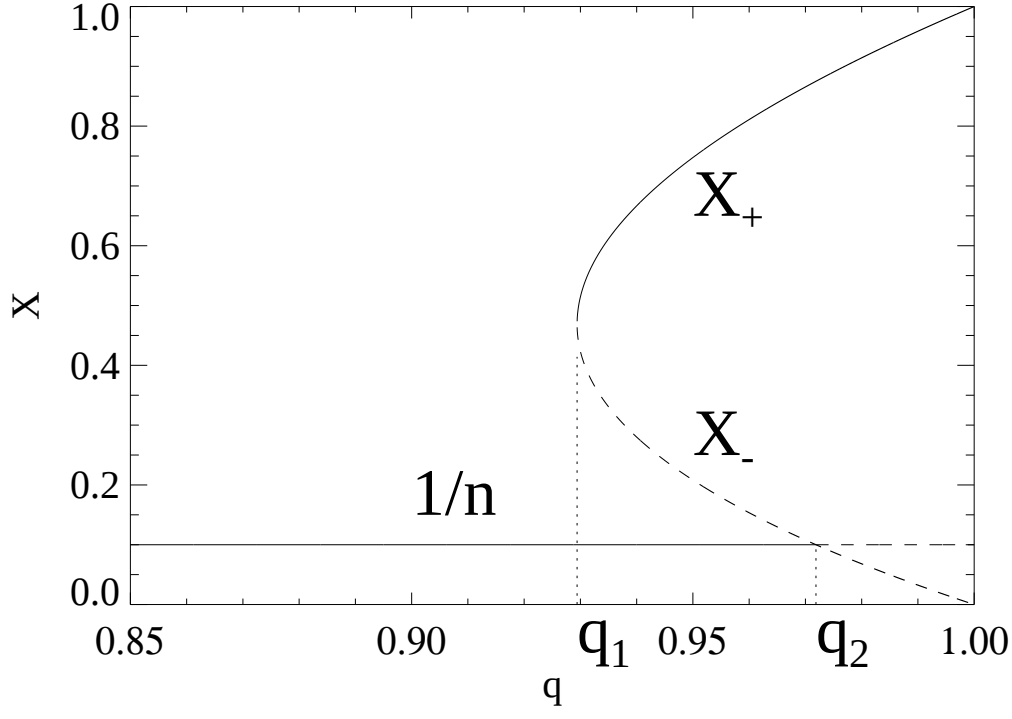
State $\{x_1, \ldots, x_n\} \in \Delta^{n-1}$

Fitness $f_i = f_0 + \sum_{j=1}^{n} F(L_i, L_j) x_j$

Dynamics $x_i(t+1) = \frac{\sum_j x_j(t) f_j Q_{ji}}{\sum_j x_j(t) f_j}$

Continuous $\dot{x}_j = \sum_i f_i x_i Q_{ij} - (\sum_i f_i x_i) x_j, \quad 1 \leq j \leq n,$

The Bifurcation



No fitness, no coherence

$$f_i = f_0$$

$$\dot{x}_i = f_0(\sum_j x_j Q_{ji}) - f_0 x_i$$

$$\mathbf{x}(t) \rightarrow (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})^T$$

Emergence of Coherence – Social Learning

1. No fitness
2. Learn from Everybody

$$\mathcal{L} = \{L_1, L_2, \ldots, L_n\}$$

$$\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$$

$$C_i \subset L_i$$

Learning Algorithm

$\mathcal{A}$: count cues k_i

$\arg\max_i \frac{k_i}{k}$ if unique

$j \sim \frac{1}{n}$ otherwise

cue i with probability ax_i

Population Dynamics

$$F : \Delta^{(n-1)} \rightarrow \Delta^{(n-1)}$$

$$F_i = f_i + \frac{1}{n}(1 - \sum_j f_j)$$

$$f_j(\mathbf{x}(t), a, k) = \sum_{\mathbf{k} \in I_j} \binom{k}{\mathbf{k}} \prod (p_j)^{k_j} \begin{cases} \mathsf{p}_i = a x_i & 1 \leq i \leq n \\ \mathsf{p}_{n+1} = 1 - a \end{cases}$$

where

$$I_j = \{(k_1, \ldots, k_{n+1}) \mid k_j \text{largest}\}$$

Bifurcations

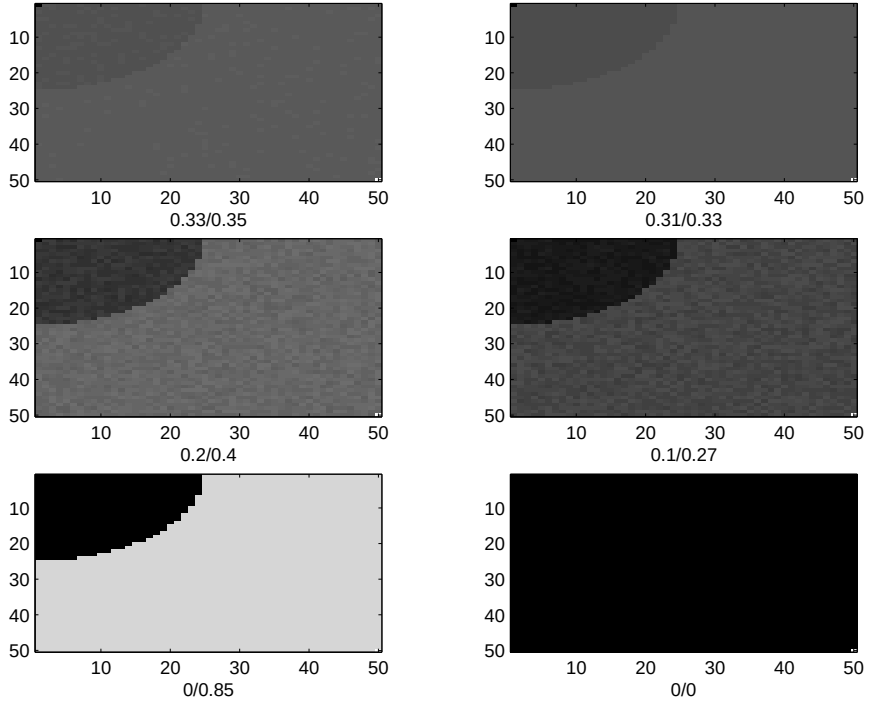
1. small k – uniform solution
2. if $k > k_a$ – new solutions
3. if $k > k_a$ – uniform solutions become unstable
4. finite number ($\leq k(2^n)$) of fixed points
5. Only n are stable.

Further Issues

1. $\mathcal{G} \equiv \mathcal{L}$: variety of linguistic theories
2. $\mathcal{A}$: variety of learning algorithms
3. $\mathcal{P}$
4. Generational structure
5. Maturation time: developmental constraints
6. Neighborhood effects and social stratification
7. Finite population effects
8. Bilingual and Multilingual settings

Spatial and Network Effects

$I(x, y)$ is influence of y on x



Major Insights

1. Different learning algorithms have different evolutionary consequences.

memoryless/batch

symmetric/asymmetric

multilingual/monlingual

single teacher/many teachers

2. Phase transition phenomena in linguistic evolution.

subtle effects of frequency

explanations of language change

explanations of dialect formation

3. Natural selection, Social connectivity, and the Emergence of Language

Coherence conditions

Social connectivity

Conclusions

1. Time scales of Language Evolution

(a) Evolutionary - origin

(b) Historical - change

2. Role of Learning

3. Role of Computational Models

4. Empirical Validation

5. There are deep connections to

(a) population models in evolutionary biology

(b) artificial life and multiagent systems

(c) bounded rationality in social sciences