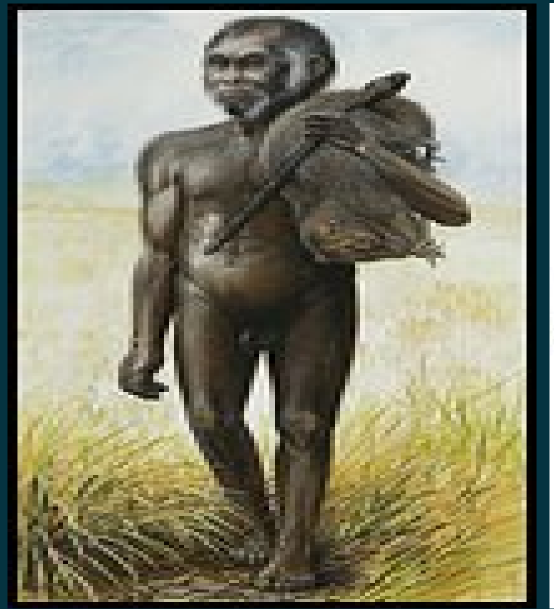
The Major Transitions in the Evolution of Language

Willem Zuidema

Language Evolution & Computation Research Unit
Institute for Cell, Animal and Population Biology
University of Edinburgh

Institute for Logic, Language and Computation
University of Amsterdam



Theories of Language Evolution

Scholar X has belief Y about how language evolved. When does X count as “scientific knowledge”?

- If Y is true.

Theories of Language Evolution

Scholar X has belief Y about how language evolved. When does Y count as "scientific knowledge"?

- If Y is true.
- If X is justified in believing Y to be true.

Theories of Language Evolution

Scholar X has belief Y about how language evolved. When does X count as “scientific knowledge”?

- If Y is true.
- If X is justified in believing Y to be true.
- If X is able to convince others of Y and its justification.

$\longrightarrow$ Empirical Testability

$\longrightarrow$ Formalisation

Evolutionary Biology

Formalisation: Population Genetics, Evolutionary Game Theory, Social Evolution Theory, Artificial Evolution *(chapter 2)*

Empirical Testability: Drosophila, sex ratios, virus evolution, co-operative and spiteful behaviour, animal communication

Syntactic Theory

Formalisation: Chomsky Hierarchy, G&B theory, HPSG, LFG, CCG, TAG, Probabilistic Parsing *(chapter 3)*

Empirical Testability: canonical syntactic phenomena, grammaticality judgments, typological data, historical data, reaction times, preferential looking, brain imaging

Why, then, is it so difficult to give a plausible and precise scenario for the evolution of language?

This talk

1. Requirements on evolutionary explanations for language

2. Previous models, and where they go wrong

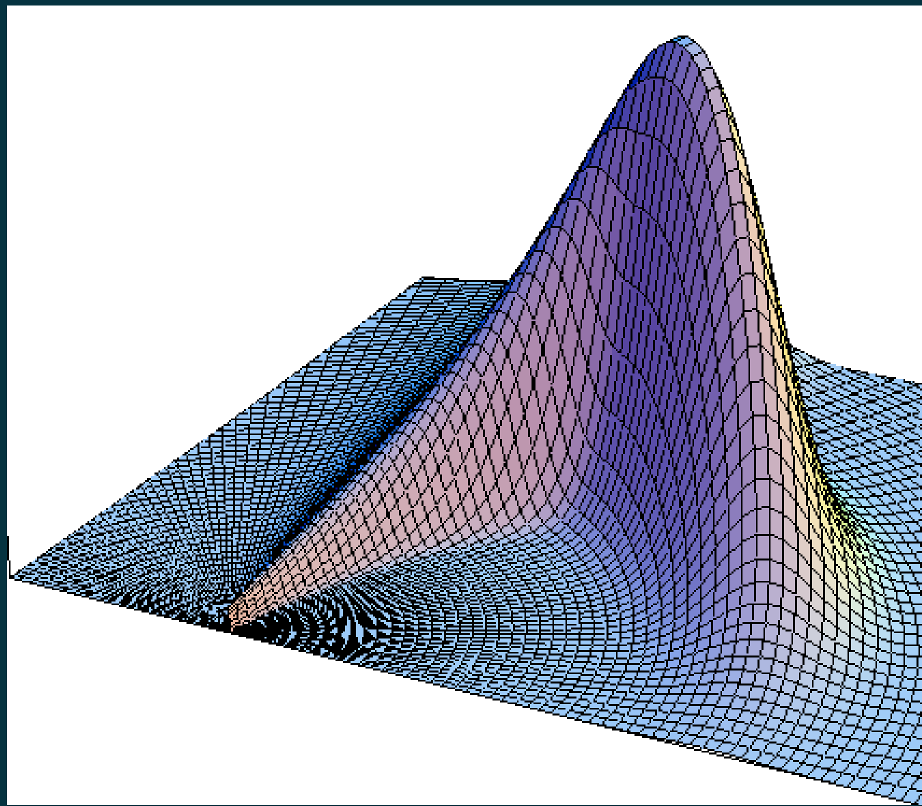
3. A new model of the evolution of phrase-structure

Structure of evolutionary explanations

Optimality Theory in Evolutionary Biology

(Parker & Maynard Smith, 1990, Nature*)*

1. Strategy set
2. Fitness function ("Payoff")
3. Paths of ever increasing fitness ("Invasibility").



Jackendoff's scenario

1. Use of symbols in a non-situation-specific fashion

↙ ↘

2. Open class

↓

3. Combinatorial Phonology ↘

4. Concatenation

↓

5. Symbol Position ↙ (Compositionality)

(Protolanguage about here)

↓

6. Hierarchical phrase-structure

↙ ↘

7. Symbols for abstract semantic relationships

8. Grammatical categories

↙ ↘

9. Inflections for semantic relationships

10. Grammatical functions for semantic relations

↘ ↓ ↙

(Jackendoff'02)

(Modern language)

Problem I: how to define a strategy set

Distinctive features? Phoneme hierarchies?

Lambda calculus?

Chomsky Hierarchy?

Principles & Parameters?

Left/right application, type raising?

Problem II: how to define a fitness function

Perceptual contrast? Articulatory ease?

Expressiveness?

Linguistic coherence?

Variation? Group identity?

Recursion?

Would it be a great advantage for one of our ancestors squatting alongside the embers, to be able to remark: "Beware of the short beast whose front hoof Bob cracked when, having forgotten his own spear back at camp, he got in a glancing blow with the dull spear he borrowed from Jack"? (David Premack, 1985).

Problem III: invasibility

(a) The Problem of Cooperation: why would senders give away (honest) information, if only the receiver benefits from it?

(Dawkins & Krebs'78; Grafen'90; Dessalles'98)

(b) The Problem of Coordination: how do senders and receivers agree on which signal to use for which meaning? How can the first user of some linguistic innovation benefit from it in a population without it? *(Lewis'69; Hurford'89)*

Proposed Solutions

(a) The Problem of Cooperation. *Solutions based population structure (kin selection, reciprocal altruism, green beards)*

(Hamilton’64; Franks’98)

(b) The Problem of Coordination. *Solutions: population structure (kin communication), “pre-adaptation” (asymmetry production & interpretation), . . .*

Jackendoff's scenario

1. Use of symbols in a non-situation-specific fashion

↙ ↘

2. Open class

↓

3. Combinatorial Phonology

↘

4. Concatenation

↓

5. Symbol Position (Compositionality)

↙

(Protolanguage about here)

↓

6. Hierarchical phrase-structure

↙ ↘

7. Symbols for abstract semantic relationships

↘

8. Grammatical categories

↙ ↘

9. Inflections for semantic relationships

↓

10. Grammatical functions for semantic relations

↙

(Modern language)

(Jackendoff'02)

Combinatorial Phonology

	Liljencrants & Lindblom'72	Lb, MacNeilage, St.-Kennedy'84	Nowak & Krakauer'99	Oudeyer'01, '02
Strategy Set	no temporal dimension	discrete trajectories	holistic + combinatorial	single strategy (continuous tr.)
Payoff Function	unrealistic for small distances	unrealistic for small distances	unrealistic (no confusion)	-
Invasibility Analysis	omitted	omitted	included	-

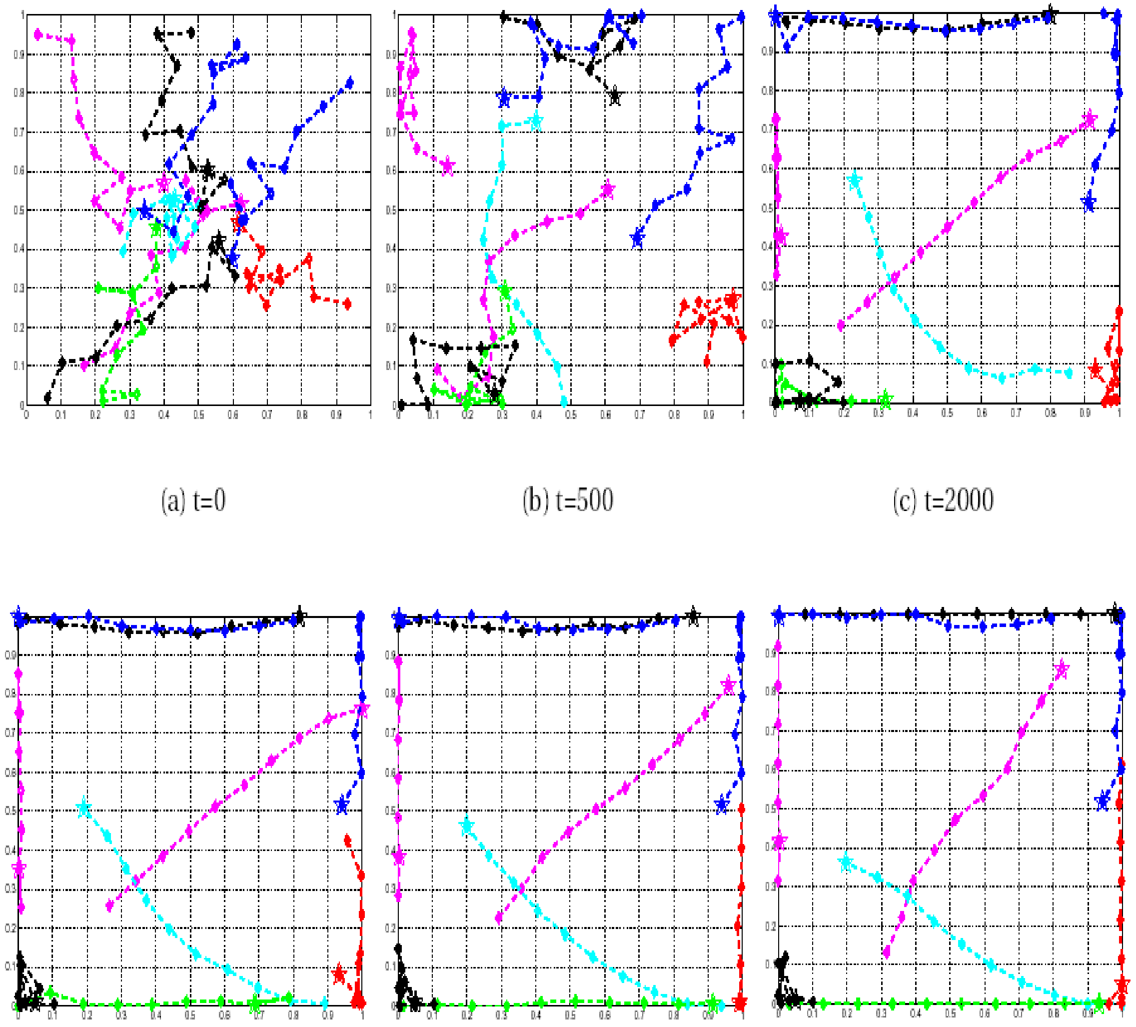
Combinatorial Phonology

(Chapter 4)

Strategy Set: continuous trajectories through acoustic space

Payoff function: based on confusion probabilities

Invasibility analysis: included



(a) t=0 (b) t=500 (c) t=2000

(d) t=4000 (e) t=8000 (f) t=13000

Compositional Semantics

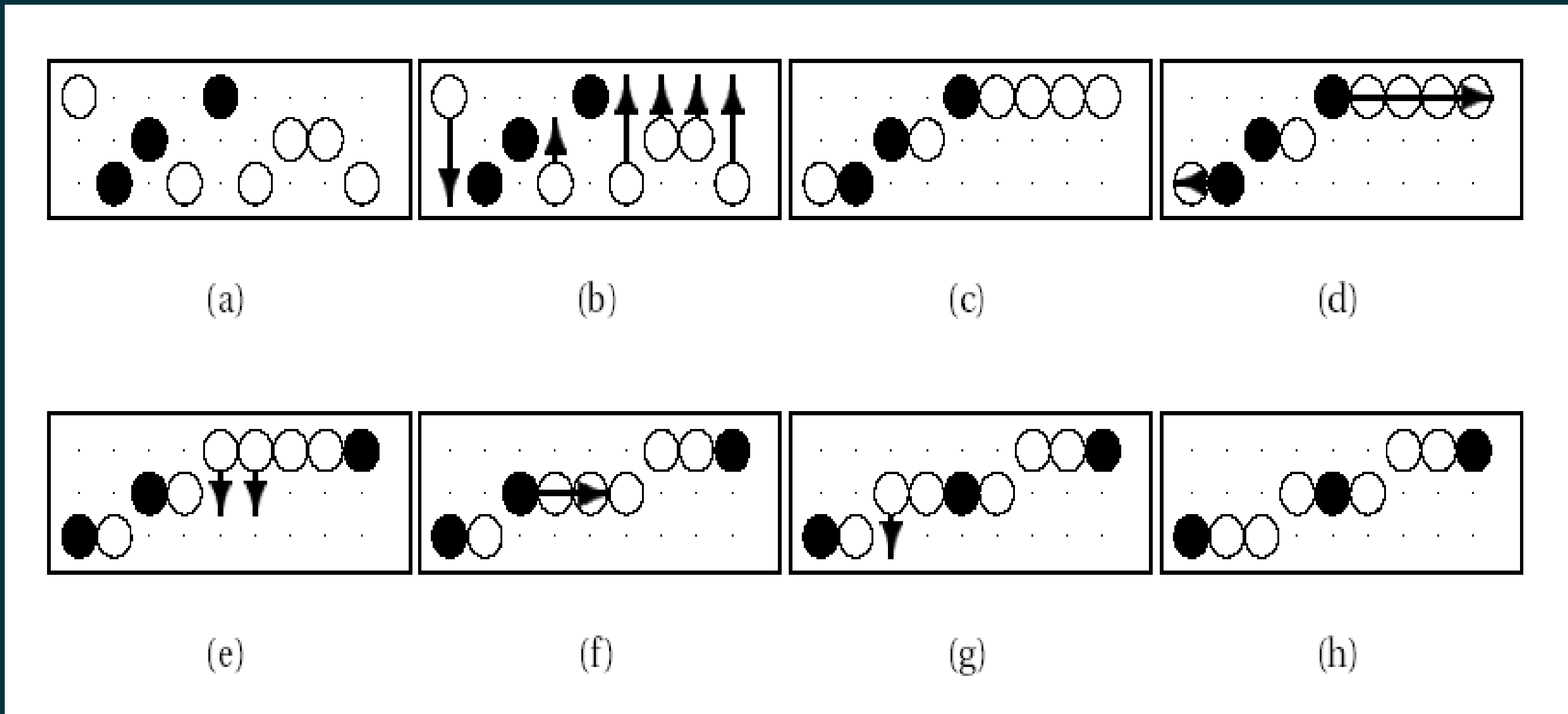
	Nowak & Krakauer'99	Nowak,Plotkin & Jansen'00	Batali'98	Kirby'00, Hurford'00 etc.
Strategy Set	holistic + compositional	holistic + compositional	single strategy	single strategy
Payoff Function	unrealistic (no confusion)	unrealistic (no confusion)	-	-
Invasibility Analysis	included	omitted	-	-

Compositional Semantics

Strategy Set: Form–meaning mappings, with similarities between forms and between meanings.

Payoff function: based on confusion probabilities

Invasibility analysis: included *(Chapter 5)*



Hierarchical Phrase-Structure

	Hashimoto & Ikegami'96	Zuidema & Hogeweg'00	Nowak,Komaro-va & Niyogi'01	Batali'02, Kirby'02 etc.
Strategy Set	Rew.Grammars (Chomsky Hier.)	Rew.Grammars (no learning)	Uniformity Assumption	single strategy (iterated learn.)
Payoff Function	freq.independent, quantitative	freq.dependent, quantitative	freq.dependent, quantitative	-
Invasibility Analysis	included	included	included	-

A new model of the evolution of syntax

(Chapter 6)

Strategy set: single learning algorithm (iterated learning)

Payoff function: frequency dependent, quantitative ($|L \cap L'|$)

Invasibility analysis: -

Compression-based Grammar Induction

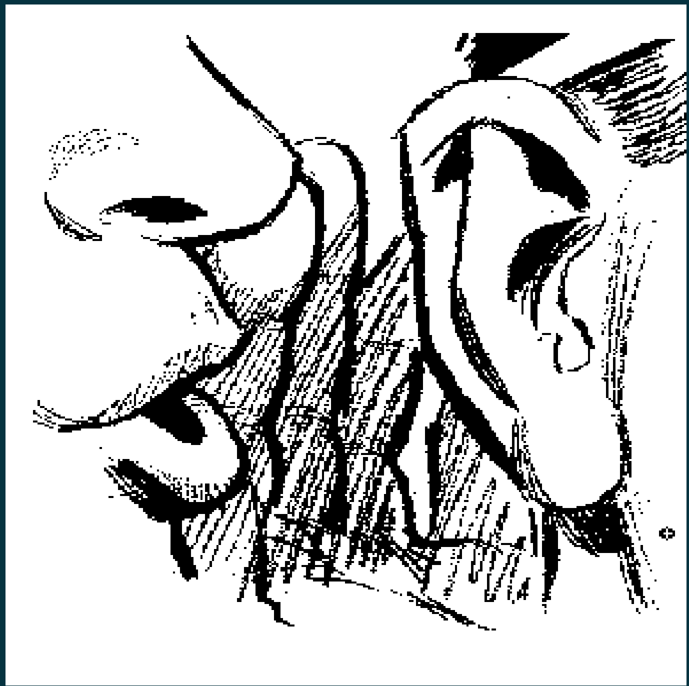
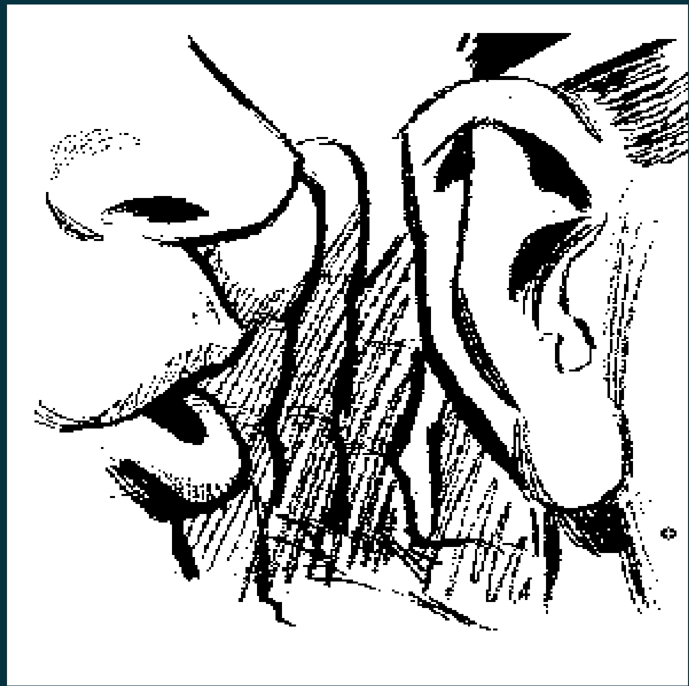
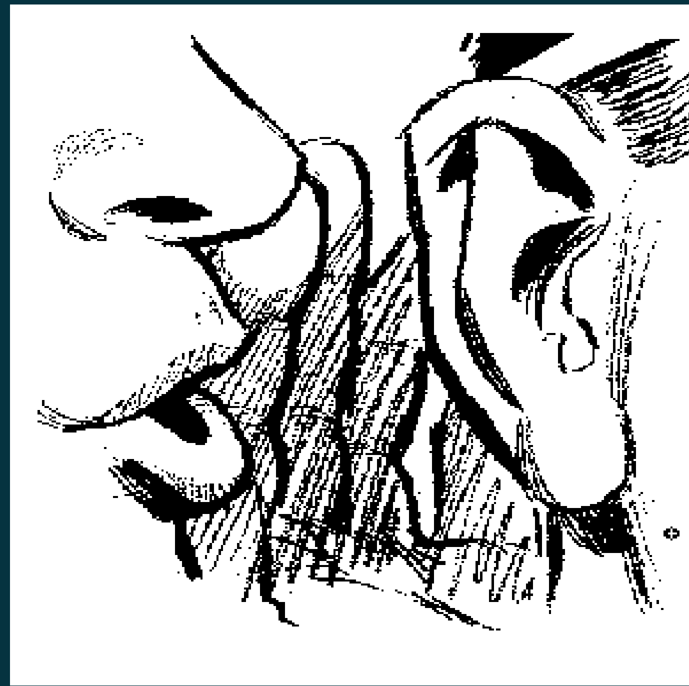
Input: example strings *(text)*, generated by target grammar G

Output: a context-free grammar G'

Three steps: incorporation, compression, generalisation

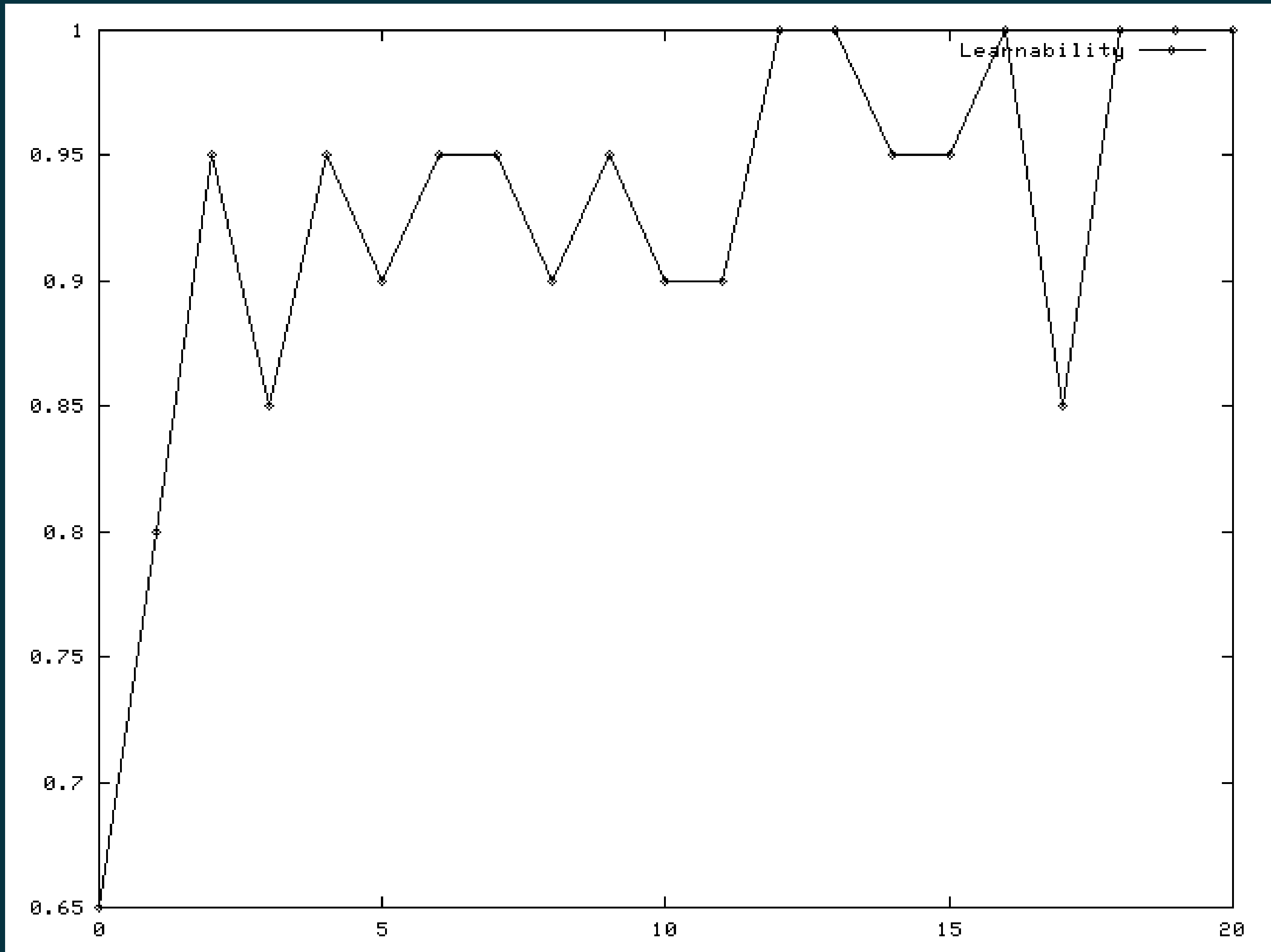
Performance measure: G' generates 100 strings, how many can G recognise?

Iterated Learning

$$
\begin{array}{ccccccccccccc}
G & \longrightarrow & L & \Longrightarrow & G' & \longrightarrow & L' & \Longrightarrow & G'' & \longrightarrow & L'' & \Longrightarrow & G''' \\
= & P & = & I & = & P & = & I & = & P & = & I & = \\
\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix} & & \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{pmatrix} & & \begin{pmatrix} r'_1 \\ r'_2 \\ \vdots \\ r'_{N'} \end{pmatrix} & & \begin{pmatrix} s'_1 \\ s'_2 \\ \vdots \\ s'_{M'} \end{pmatrix} & & \begin{pmatrix} r''_1 \\ r''_2 \\ \vdots \\ r''_{N''} \end{pmatrix} & & \begin{pmatrix} s''_1 \\ s''_2 \\ \vdots \\ s''_{M''} \end{pmatrix} & & \begin{pmatrix} r'''_1 \\ r'''_2 \\ \vdots \\ r'''_{N'''} \end{pmatrix}
\end{array}
$$

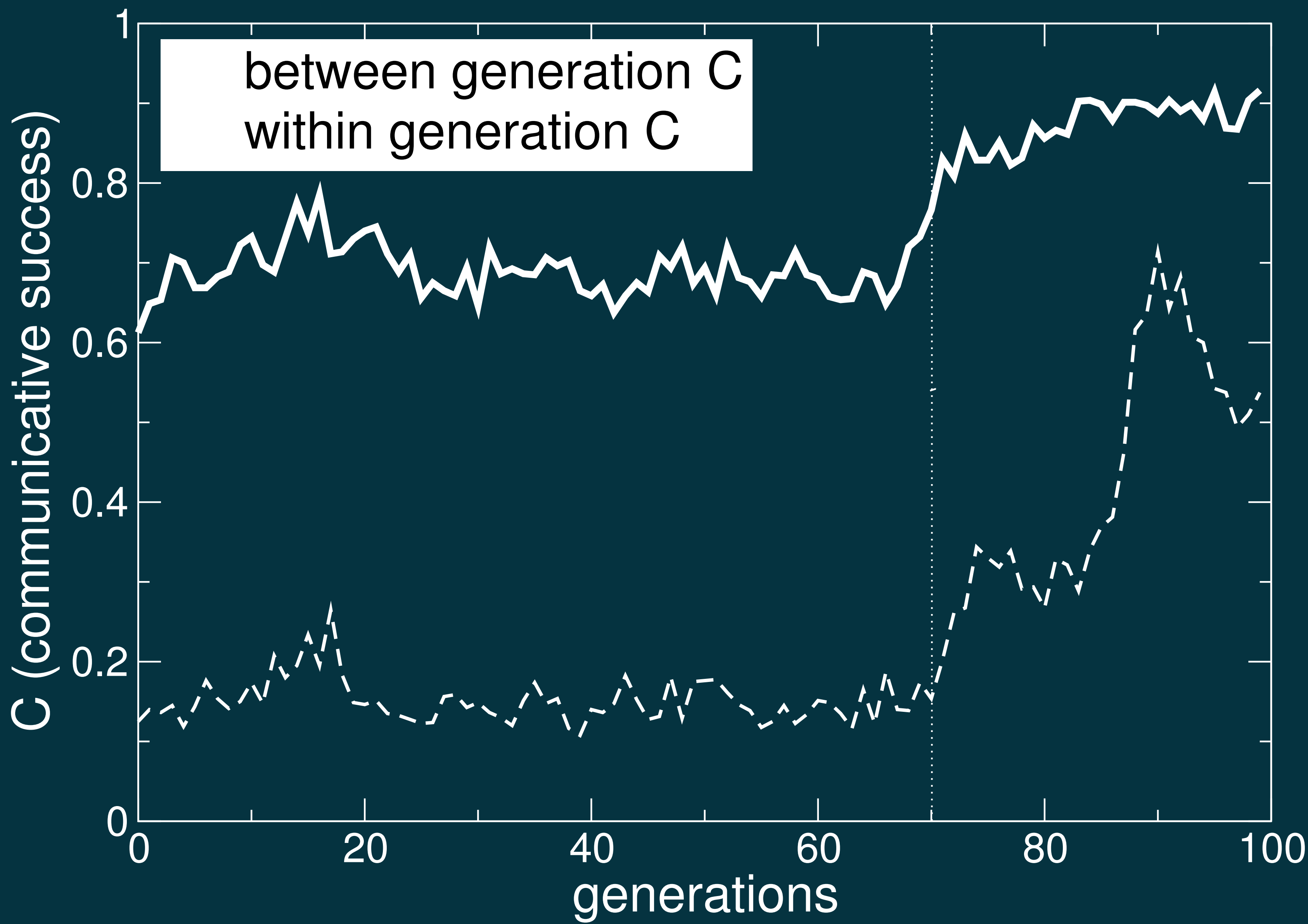
The emergence of learnability



Transmission

Iterated learning: individuals in a *chain* learn from the previous individual and teach to the next;

Fitness proportional selection: the fitness of an individual is determined by its success in communicating with the individuals of its own generation. The expected number of offspring is proportional to this fitness.



Parameters: hearer benefit condition, $V_t = \{0, 1, 2, 3\}$, $V_{nt} = \{S, a, b, c, d, e, f\}$, P=20, T=100, M=100, d=8, l_0=12

The relevance of cultural evolution for biological evolution

- The “appearance of design” cannot be taken as evidence for biological adaption;
- A biased algorithm can do better than an unbiased algorithm; hence, lower bounds derived for an unbiased algorithm *(Nowak, Komarova & Niyogi’01, Science)* are not valid.
- The payoff function over learning mechanisms needs to take into account the effects they have on the languages spoken in a population.

Conclusions: Requirements on theories of language evolution

- be formal,
- be testable,
- specify a reasonable strategy set,
- specify a reasonable payoff function,
- show that each innovation can invade and get established.

To meet these criteria, I have presented three new models of the evolution of combinatorial phonology, compositional semantics and hierarchical phrase-structure.

Why is it so difficult to give a plausible and precise scenario for the evolution of language?

Linguistics has focused on mentalist models, that abstract away from the actual (neural) instantiations of the abstract computations they postulate, the function of language in daily life and the way knowledge of language is acquired and transmitted. Yet it is these things – the strategy set and the payoff function – that evolutionary explanations are concerned with.

Acknowledgements

Supervisors: Simon Kirby, LEC, Linguistics
Jim Hurford, LEC, Linguistics
Nick Barton, Institute of Evolutionary Biology

Collaborators: Andy Gardner, Kingston, Canada (ch. 2)
Bart de Boer, Groningen, the Netherlands (ch. 4)
Gert Westermann, Oxford Brookes, U.K. (ch. 5)
Tim O’Donnell, Harvard, USA (ch. 6,7)

Colleagues at the VUB AI-Lab, LEC, ICAPB & ILLC

The structure of these models

(Nowak & Krakauer'99; my chapter 4,5)

$$F(L, L) < F(L', L) < F(L', L').$$

- L is the resident language; L' is an innovation (mutant);
- $F(L, L')$ is the payoff of a resident speaking with a mutant.

Possibly, $F(L, L) > F(L''''', L)$, but still:

$F(L, L) < F(L', L) < F(L', L') < F(L'', L') < \ldots < F(L''''', L''''')$.

What we really want to show

(Harley'81; Maynard Smith'82; Hurford'89; K. Smith'04; my chapter 6)

$$F(A(L), A(L)) < F(A'(L), A(L)) < F(A'(L'), A'(L')).$$

- A is the resident learning algorithm; A' is the mutant;
- $A(L)$ is language that results from learning with A from L;
- More mutants in the population will change the population's language from L to L'.

Representation

Context-free grammars : $A \mapsto t$, $A \mapsto BC$, $A \mapsto Bt$.

Parsing: depth-first search with maximum depth d

Derivation: random string from parsable language

Interaction: speaker derives random string s, and hearer checks if s is a string from its own language. If so, the interaction is a success, otherwise it is a failure.

Grammar Induction

Incorporation: extend the language, such that it includes the encountered string

Compression: substitute frequent and long substrings with a nonterminal

Generalization: equate two nonterminals if they occur frequently in the same context

off-line learning, sufficient expressiveness enforced

Example: the evolution of alarm calls

“grunt” “chirp”

Payoff matrices

sender sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$-c$	$-c$	$-c$
{grunt, chirp}	$-c$	$-c$	$-c$
silent	0	0	0

receiver sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$+b$	$-m$	0
{grunt, chirp}	$-m$	$+b$	0
silent	$-w$	$-w$	0

Evolutionary Stable Strategies

sender sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$-c$	$-c$	$-c$
{grunt, chirp}	$-c$	$-c$	$-c$
silent	0	0	0

receiver sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$+b$	$-m$	0
{grunt, chirp}	$-m$	$+b$	0
silent	$-w$	$-w$	0

Evolutionary Stable Strategies

sender sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$-c$	$-c$	$-c$
{grunt, chirp}	$-c$	$-c$	$-c$
silent	0	0	0

receiver sender strategy ↓	receiver strategy {leopard, eagle}	 {eagle, leopard}	 ignore
{chirp, grunt}	$+b$	$-m$	0
{grunt, chirp}	$-m$	$+b$	0
silent	$-w$	$-w$	0

Language Evolution = climbing the Chomsky Hierarchy?

Finite $\longrightarrow$ Finite State $\longrightarrow$ Context Free $\longrightarrow$ Mildly Context-Sensitive?