Gene-Language Coevolution

Ted Briscoe

Computer Laboratory
Natural Language and Information Processing Group
University of Cambridge

ENS, Paris
Mar 2014

Genetic assimilation (Waddington)

Fruit flies and heat $\rightsquigarrow$ cross veined wings:

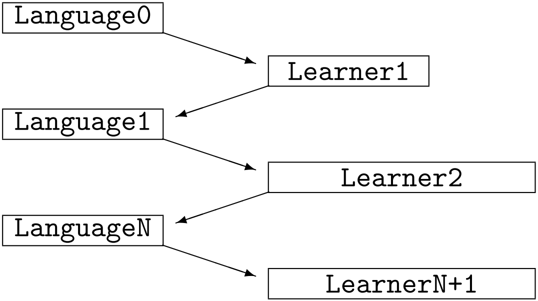
If there were selection for the ability to use language, then there would be selection for the capacity to acquire the use of language, in interaction with a language-using environment; and the result of selection for epigenetic responses can be, as we have seen, a gradual accumulation of so many genes with effects tending in this direction that the character gradually becomes genetically assimilated (*The Evolution of an Evolutionist*, p305-6)

- "Gene-culture co-evolution" (Durham) – lactose tolerance
- "Baldwinian niche construction" (Deacon) – (un)masking
- "Evolutionary Bayesianism" (Geisler & Diehl) — priors evolve

Evolutionary Bayesian Learning

Repeated relearning of languages, based on the output of the previous generation, by learners with evolving learning biases

Mutate the prior probabilities of parameters / constraints in the learner



The Baldwin Effect (Pinker & Bloom)

- A (proto)language emerges
- Learning Cost – unsuccessful communication (fitness)
- Natural Selection – for individuals who learn better
- Gradual Coevolutionary Adaptation, not Saltation
- Linguistic Universals – evidence for GA/BE?
- But universals are result of (convergent) linguistic evolution

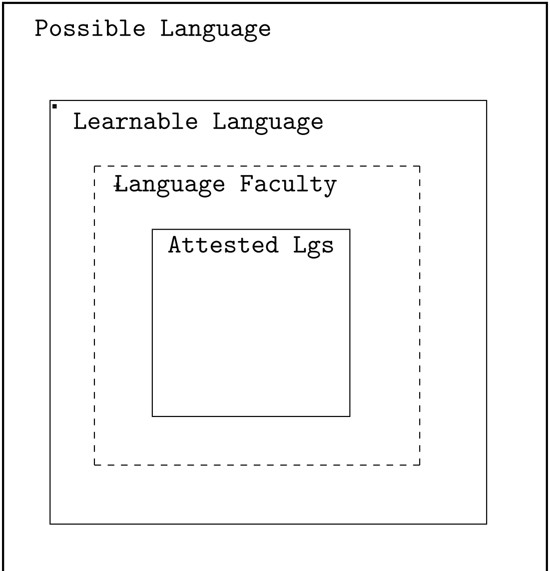
Deacon – no GA/BE

Languages don't just change they *evolve*. And children themselves are the rigged game. Languages are under powerful selection pressure to fit children's likely guesses, because children are the vehicle by which a language gets reproduced. Languages have to adapt to children's spontaneous assumptions... because children are the only game in town. ... languages need children more than children need languages. (Terry Deacon, *The Symbolic Species*, 1997:109)

Deacon – (Un)Masking

- Genetic assimilation requires environmental constancy
- (Proto)language unmasked (sequential/symbolic) learning
- Linguistic evolution much faster then biological evolution
- No GA/BE – languages evolved to be learnable
- Predicts close fit between cognition and language
- Solves 'problem' of FLA / 'Poverty of the Stimulus' (Zuidema)
- But language change may not cover the entire hypothesis space

Deacon vs. Pinker & Bloom



Coevolutionary Stochastic ILM

- Language Agent: $(LAgt_i)$
  $lg^j = LP(UG, fm_k), m_k = Parse(lg^j, f_k),$
  $f_k = Generate(lg^j, m_k), Age(0:9), Fit(0:1) >$
- Successful Interaction:
  $(LAgt_i, LAgt_j), i \neq j,$
  $f_k = Generate(lg^i, m_k), m_k = Parse(lg^j, f_k)$
- Reproduction: $(\propto Fit)$:
  $(LAgt_i, LAgt_j), i \neq j, Age_{i/j} > 3,$
  $Mutate(Crossover(Prior(LAgt_i, Prior(LAgt_j))))$
- Migrations: $\{LAgt_1, \ldots LAgt_m + LAgt_{m+1}, \ldots LAgt_n\}$
  $LAgt_{m+i} :< lg \neq lg^d, Age(> 3) >$

Coevolutionary Stochastic ILM

- Language Agent: $(LAgt_i)$
  $lg^j = LP(UG, fm_k), m_k = Parse(lg^j, f_k),$
  $f_k = Generate(lg^j, m_k), Age(0:9), Fit(0:1) >$
- Successful Interaction:
  $(LAgt_i, LAgt_j), i \neq j,$
  $f_k = Generate(lg^i, m_k), m_k = Parse(lg^j, f_k)$
- Reproduction: $(\propto Fit)$:
  $(LAgt_i, LAgt_j), i \neq j, Age_{i/j} > 3,$
  $Mutate(Crossover(Prior(LAgt_i, Prior(LAgt_j))))$
- Migrations: $\{LAgt_1, \ldots LAgt_m + LAgt_{m+1}, \ldots LAgt_n\}$
  $LAgt_{m+i} :< lg \neq lg^d, Age(>3) >$

Coevolutionary Stochastic ILM

- Language Agent: $(LAgt_i)$
  $lg^j = LP(UG, fm_k), m_k = Parse(lg^j, f_k),$
  $f_k = Generate(lg^j, m_k), Age(0:9), Fit(0:1) >$
- Successful Interaction:
  $(LAgt_i, LAgt_j), i \neq j,$
  $f_k = Generate(lg^i, m_k), m_k = Parse(lg^j, f_k)$
- Reproduction: $(\propto Fit)$:
  $(LAgt_i, LAgt_j), i \neq j, Age_{i/j} > 3,$
  $Mutate(Crossover(Prior(LAgt_i, Prior(LAgt_j))))$
- Migrations: $\{LAgt_1, \ldots LAgt_m + LAgt_{m+1}, \ldots LAgt_n\}$
  $LAgt_{m+i} :< lg \neq lg^d, Age(> 3) >$

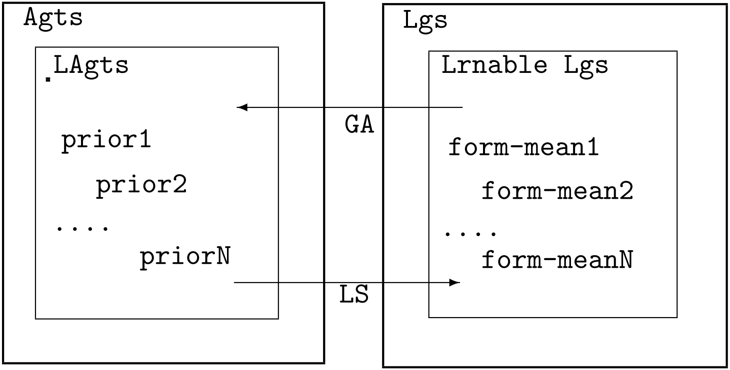
Coevolutionary Stochastic ILM

- Language Agent: $(LAgt_i)$
  $lg^j = LP(UG, fm_k), m_k = Parse(lg^j, f_k),$
  $f_k = Generate(lg^j, m_k), Age(0:9), Fit(0:1) >$
- Successful Interaction:
  $(LAgt_i, LAgt_j), i \neq j,$
  $f_k = Generate(lg^i, m_k), m_k = Parse(lg^j, f_k)$
- Reproduction: $(\propto Fit)$:
  $(LAgt_i, LAgt_j), i \neq j, Age_{i/j} > 3,$
  $Mutate(Crossover(Prior(LAgt_i, Prior(LAgt_j))))$
- Migrations: $\{LAgt_1, \ldots LAgt_m + LAgt_{m+1}, \ldots LAgt_n\}$
  $LAgt_{m+i} :< lg \neq lg^d, Age(> 3) >$

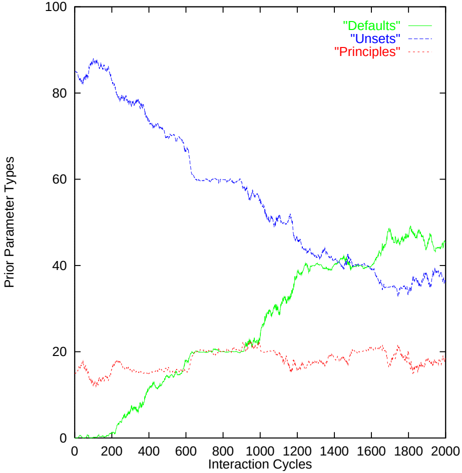
LAgt Fitness

1. Generate cost: 1 (GC)
2. Parse cost: 1 (PC)
3. Success benefit: 1 (SI)
4. Fitness function: $\frac{SI}{GC+PC}$
5. Indirect selection for expressivity – learning a partial grammar will impact on successful interactions
6. Mutation +/-1 on single prior numerator/denominator with correction for neutral, unset (p=0.5) priors so genuinely 'random'

Coevolution of Languages and Language Faculties



Assimilation of Soft Biases



Simulation Assumptions / Results

- Communicative success confers fitness on its users (larynx/choking)
- Lg change about an order of magnitude faster than the fastest genetic change (i.e. no. of int. cycles to fixation)
- Speed limit to lg change – >90% succ. int. in a speech community
- Population / Grammar Size – mean 5% of grammar space explored in time it takes a mutation to go to fixation in population
- Genetic assimilation asymptotes in the face of lg change – no prediction of genetic fixation on one grammar
- Soft biases are preferred to hard constraints

Gene-Sign (De)Correlation

- P-setting encoding of prior/posterior parameter values
- Phenotype (A Grammar) and Genotype ($LP(UG)$) correlate
- Epistasis and Pleotropy suggest not realistic
- Decorrelate P-setting via *Mutate*
- Locality mutate more settings per mutation
- Degree alter priors more – random stronger lrng biases

(De)Correlation Results

1. More learners fail to acquire (full) grammars (1–24%)
2. More premptive mutations cause linguistic change (2–99%)
3. More speech communities converge to subset grammars (5–100%)
4. More language change without migration (4-84%)
5. Complex adaptive systems poised at 'edge of chaos' (Kaufmann) – the evolution of evolvability (slight decorrelation is optimal)

The Logical Problem of Lg Evolution

- How did UG emerge? (Chater, Christiansen, et al)
- UG is arbitrary (not functional) (why?!)
- UG not by saltation – prob 2^{-2500} (Pinker & Bloom)
- UG by gradual evolution, but fast lg. change (Deacon)
- Lg emerge 100KYA, people dispersed, different UGs? – need to track fast change! (Dediu – tone)
- Brain shapes lg. = neural network, processing limits, etc
- Poor simulation model – no communicative success, etc.

Language-specific Learning Biases?

- Creolisation: SVO word order, Tense-Aspect system (Bickerton)
- Atomisation: not / ne..pas lex./clitic/morph. negation (Wanner/Gleitman)
- Linear sequencing: rolling down manner/path motion, serial verbs(NSL, Senghas)
- Abstract parameters of variation: Pro-drop – Italian/English Old/Mod. French, 2 possibilities not 64 (Baker)
- Overregularisation / errors restricted: goed / falled my dolly down / did you saw (see) it but not: did the man who saw? the did man see

Timescales, Sizes, and Speed Limits

- Speed limit to linguistic evolution – successful interactions must predominate in a speech community: SI $> 90\%$, language contact, networks of interaction, etc?
- Speed limit to biological evolution – phenotypes must function even under strong selection: 1 bit/generation (Worden), N bits/generation (Mackay), $N =$ population size, (10K out of Africa)
- When did (proto)language emerge? 2.5M (Deacon) – 50K (Chomsky) years ago: 2.5KB – 20MBs?!
- FOXP2, lactose, etc – 10% of human genome affected by selection in past 50K years (but phenotypic effects?)
- Even at lower bound, room for modest GA on priors, and generic inductive bias (Occam's Razor) is present either way

Summary

- Evolutionary Bayesianism – good model for examining how inductive bias might have evolved in linguistic evolutionary niche
- Deacon's argument that languages change too fast for GA wrong so long as some part of the hypothesis space is not manifest during time mutation to go to fixation
- Biased parameters rather than principles or single grammars will emerge by GA given language change
- GA is real and there has been enough time since (proto)language emerged
- (Domain-specific) inductive bias for 'natural grammars' might explain e.g. commonalities amongst creoles

Reading

Geisler, W. & Diehl, R. “A Bayesian approach to the evolution of perceptual and cognitive systems” *Cognitive Science*, 27 (2003).

Weber, B. & Depew, D., *Evolution and Learning: The Baldwin Effect Reconsidered* MIT 2003

Briscoe, E. “Grammatical Assimilation” in Kirby, S & Christiansen, M., *Language Evolution*, OUP, 2003

www.cl.cam.ac.uk/users/ejb/

Zuidema, W., “How the poverty of the stimulus solves the poverty of the stimulus”

Christiansen, M, Chater, N., et al “Language as shaped by the brain”

www.langev.com