An Evolutionary Linguistic Model

Ted Briscoe

Computer Laboratory
Natural Language and Information Processing Group
University of Cambridge

ENS, Paris
Mar 2014

An Analogy

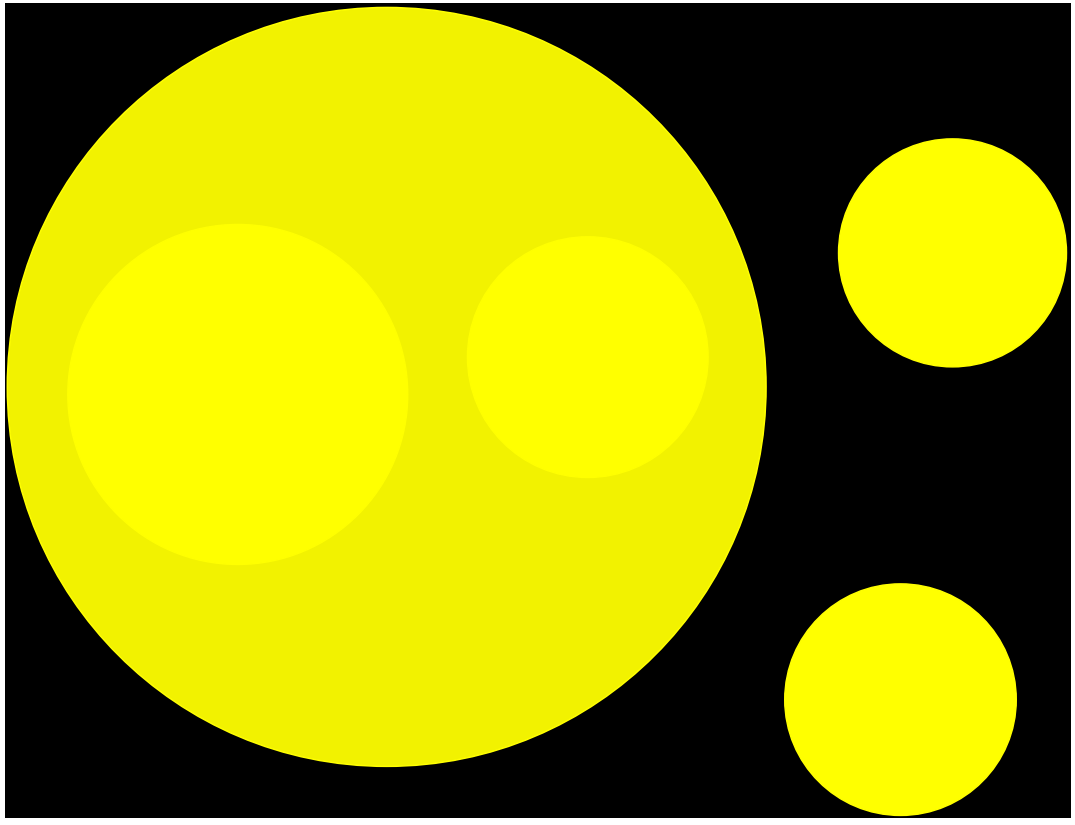
A drunk had lost his keys on the street and was frantically searching for them under a streetlamp. ‘Where did you drop them?’ asked a concerned passer by. ‘Over there’ he replied, indicating a spot 30 yards away. ‘So why are you looking here under the lamp?’ ‘The light is better here’.

Epistemology / Philosophy of Science

Karl Popper:
No logic of discovery
Logic of justification (methodological falsification)

Kantian Spectacles: We interpret and attempt to explain data in terms of our favourite theories / intellectual training

Hypothesis Space(s)



How do we weight the contribution of different factors / theories?

Linguistic Universals

- Compositionality: Vervet monkeys vs. Bees
- MorphoSyntax: Subj-Vb-Obj, Subj-Obj-Vb, Case: he/him
- Mild Context Sensitivity: nesting (a^n b^n, aabb), cross-serial (a^n b^n c^n, aabbcc), intersecting (a^n,b^n,c^n, cabbca)
  - The guy Kim kissed smiled
  - Kim den huus helped paint
- Productivity / Recursion: Bees vs. Akkadian / Piraha vs. English

Linguistic Universals

- Compositionality: Vervet monkeys vs. Bees
- MorphoSyntax: Subj-Vb-Obj, Subj-Obj-Vb, Case: he/him
- Mild Context Sensitivity: nesting (a^n b^n, aabb), cross-serial (a^n b^n c^n, aabbcc), intersecting (a^n,b^n,c^n, cabbca)
  - The guy Kim kissed smiled
  - Kim den huus helped paint
- Productivity / Recursion: Bees vs. Akkadian / Piraha vs. English

Linguistic Universals

- Compositionality: Vervet monkeys vs. Bees
- MorphoSyntax: Subj-Vb-Obj, Subj-Obj-Vb, Case: he/him
- Mild Context Sensitivity: nesting (a^n b^n, aabb), cross-serial (a^n b^n c^n, aabbcc), intersecting (a^n,b^n,c^n, cabbca)
  - The guy Kim kissed smiled
  - Kim den huus helped paint
- Productivity / Recursion: Bees vs. Akkadian / Piraha vs. English

Linguistic Universals

- Compositionality: Vervet monkeys vs. Bees
- MorphoSyntax: Subj-Vb-Obj, Subj-Obj-Vb, Case: he/him
- Mild Context Sensitivity: nesting (a^n b^n, aabb), cross-serial (a^n b^n c^n, aabbcc), intersecting (a^n,b^n,c^n, cabbca)
  - The guy Kim kissed smiled
  - Kim den huus helped paint
- Productivity / Recursion: Bees vs. Akkadian / Piraha vs. English

Linguistic Variation

- Vocabulary: size, specific form:meaning associations
- Syntax: case marking vs. word order, categories, recursion
- Change: new words, pronunciations, lg genesis (Creolisation, Nicaraguan Sign Language)

Linguistic Variation

- Vocabulary: size, specific form:meaning associations
- Syntax: case marking vs. word order, categories, recursion
- Change: new words, pronunciations, lg genesis (Creolisation, Nicaraguan Sign Language)

Linguistic Variation

- Vocabulary: size, specific form:meaning associations
- Syntax: case marking vs. word order, categories, recursion
- Change: new words, pronunciations, lg genesis (Creolisation, Nicaraguan Sign Language)

Statistical Tendencies

- Word Order: SVO, SOV $>$ VSO $>$ VOS $>$ OVS, OSV
- Correlations: OV $\rightsquigarrow$ Rel+N $\wedge$ Case
- Change: S-shaped (logistic) diffusion, grammaticalisation (delexification)
- Irregularity / Frequency: irregular forms are frequent
- Morphology / 2nd Lg Users: isolating lgs have higher proportions of 2nd lg learners/users

Statistical Tendencies

- Word Order: SVO, SOV $>$ VSO $>$ VOS $>$ OVS, OSV
- Correlations: OV $\rightsquigarrow$ Rel+N $\wedge$ Case
- Change: S-shaped (logistic) diffusion, grammaticalisation (delexification)
- Irregularity / Frequency: irregular forms are frequent
- Morphology / 2nd Lg Users: isolating lgs have higher proportions of 2nd lg learners/users

Statistical Tendencies

- Word Order: SVO, SOV $>$ VSO $>$ VOS $>$ OVS, OSV
- Correlations: OV $\rightsquigarrow$ Rel+N $\wedge$ Case
- Change: S-shaped (logistic) diffusion, grammaticalisation (delexification)
- Irregularity / Frequency: irregular forms are frequent
- Morphology / 2nd Lg Users: isolating lgs have higher proportions of 2nd lg learners/users

Statistical Tendencies

- Word Order: SVO, SOV $>$ VSO $>$ VOS $>$ OVS, OSV
- Correlations: OV $\rightsquigarrow$ Rel+N $\wedge$ Case
- Change: S-shaped (logistic) diffusion, grammaticalisation (delexification)
- Irregularity / Frequency: irregular forms are frequent
- Morphology / 2nd Lg Users: isolating lgs have higher proportions of 2nd lg learners/users

Statistical Tendencies

- Word Order: SVO, SOV $>$ VSO $>$ VOS $>$ OVS, OSV
- Correlations: OV $\rightsquigarrow$ Rel+N $\wedge$ Case
- Change: S-shaped (logistic) diffusion, grammaticalisation (delexification)
- Irregularity / Frequency: irregular forms are frequent
- Morphology / 2nd Lg Users: isolating lgs have higher proportions of 2nd lg learners/users

Origins / Protolanguage / Distribution

- Why did language emerge? (basic function / honest signals)
- Was protolanguage ever holistic / gestural?
- How many lgs/dialects/idiolects have there ever been?
- Why are some lgs spoken by more people than others?

Origins / Protolanguage / Distribution

- Why did language emerge? (basic function / honest signals)
- Was protolanguage ever holistic / gestural?
- How many lgs/dialects/idiolects have there ever been?
- Why are some lgs spoken by more people than others?

Origins / Protolanguage / Distribution

- Why did language emerge? (basic function / honest signals)
- Was protolanguage ever holistic / gestural?
- How many lgs/dialects/idiolects have there ever been?
- Why are some lgs spoken by more people than others?

Origins / Protolanguage / Distribution

- Why did language emerge? (basic function / honest signals)
- Was protolanguage ever holistic / gestural?
- How many lgs/dialects/idiolects have there ever been?
- Why are some lgs spoken by more people than others?

Basic Generative Tenets

1. A sentence is a string, a language is a stringset
2. A grammar is a finite-set of rules which defines all and only the grammatical strings of a language and their structural descriptions
3. A grammar (in the limit) defines the language of a single speaker at one instant in time
4. Universal grammar is a set of constraints on grammar rules which expresses commonalities amongst human languages

Grammars Relate Meaning to Form

Bar-Hillel '53; Chomsky '65; Montague '70; Steedman '00

Linguistic theory is concerned with an ideal speaker-listener in a completetely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, errors (random or characterisitc) in applying his knowledge of language in actual performance. (*Aspects of the Theory of Syntax*, p3–4)

A fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how this sentence is understood by the ideal speaker-hearer. (*Aspects...*, p4–5)

A More Recent Formulation

- Universal Grammar (UG) is a set of constraints on grammars which capture commonalities amongst human languages
- A grammar is a finite set of constraints which defines all and only the grammatical strings of a language and SF-LF mappings
- A grammar (in the limit) defines the I-language of a single speaker at a single instant in time (idiolect)
- Grammatical acquisition is parametric (v2 on/off, head-initial/final) finite set of finite-valued parameters
- (I-)Language change is a result of parameter resetting ('reanalysis') across generations ('immediate')
- Diffusion of change E-language across a speech community is slower (S-shaped, Kroch)
- E-Languages are dynamical systems – the aggregate output of a changing population of speakers

Not Biolinguistics!

- Emergence of the (narrow) language faculty
- Internal/External Merge = recursion
- + 'Third Factors' = Cognitive Constraints
- Exaption / Spandrels – lg. as a side-effect
- Saltation / Macromutation – no natural selection
- No account of (E-)language (change)

Not Origins!

- Language emerged 2.5M-50K years ago – no fossils!
- Brain size (Deacon/Johansson) vs. Tool Use / Culture (Everyone else)
- Skull Shape (Descended Larynx), 1M years ago?
- Irrelevant (almost) to (E-)language (change)
- Evolution is as much about maintenance and development as emergence of a trait

Universal Darwinism

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)

Evolution is Modality Free

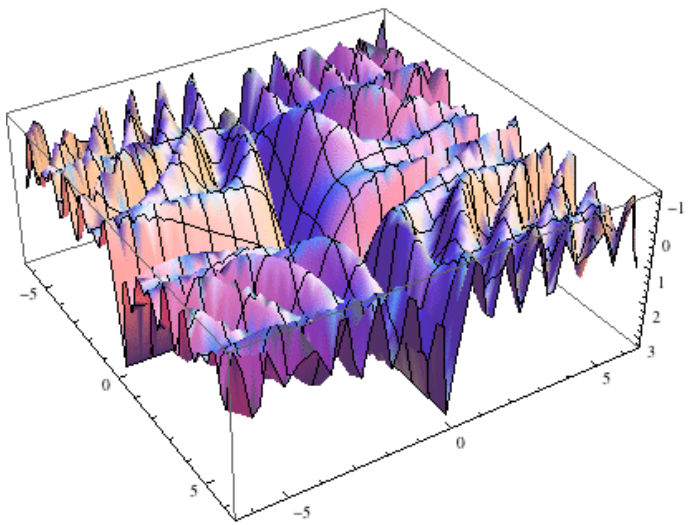
1. Variation +
2. Inheritance +
3. Selection/Drift =
4. Evolution

1. Linguistic Variation +
2. 1st/2nd Language Acquisition/Lifetime Learning +
3. Linguistic Selection/Drift =
4. Linguistic Evolution

Linguistic Selection

1. Learnability – frequency, interpretability, learning bias...
2. Expressivity – economy, memorability, prestige...
3. Interpretability – working memory, distance, (un)ambiguity...

Languages are complex adaptive systems – Multipeaked and dynamic fitness landscapes:



Nature of Evolutionary Explanation

- Irreducibly historical / diachronic (Haspelmath), possibly contingent e.g. colonisation & language
- The Invisible Hand / Self-Organisation (Keller) – local idiolectal change (I-lg) leads to global change (E-lg)
- Universals / Commonalities = Convergent Evolution – eyes, fins...
- Not just non-teleological functionalism (Givon) via adaptation
- Random Drift (Kimura) vs. Selected Change – evolution uses up variation
- Frequency-dependent selection/drift = positive feedback effects which speed up evolution – infectious diseases / trendy words

Centrality of Learning

The best linguistic 'replicators' will be frequently expressed in learner input, compatible with learning biases, easy to produce, perceive... a trade-off

Regularities (linguistic generalizations) compatible with learning biases (representable) will be more frequently expressed in learner input than ir/subregularities. Therefore, irregularity will be linked to high frequency forms and have other advantages (ease of production) to survive

Linguistic evolution (language change) = 'blind' local moves in a dynamic and complex adaptive landscape with many local optima (tinkering rather than centralised planning)

Generalized Categorial Grammar (Steedman / Lambek)

Forward/Backward Application (F/B A):

$X|Y\ Y \Rightarrow X$ $\qquad\qquad$ $\lambda\ y\ [X(y)]\ (y) \Rightarrow X(y)$

Forward/Backward/Mixed Composition (F/B/M C):

$X|Y\ Y|Z \Rightarrow X|Z$ $\qquad\qquad$ $\lambda\ y\ [X(y)]\ \lambda\ z\ [Y(z)] \Rightarrow\ \lambda\ z\ [X(Y(z))]$

Lexical/Derivational (Generalized Weak) Permutation (L/D P):

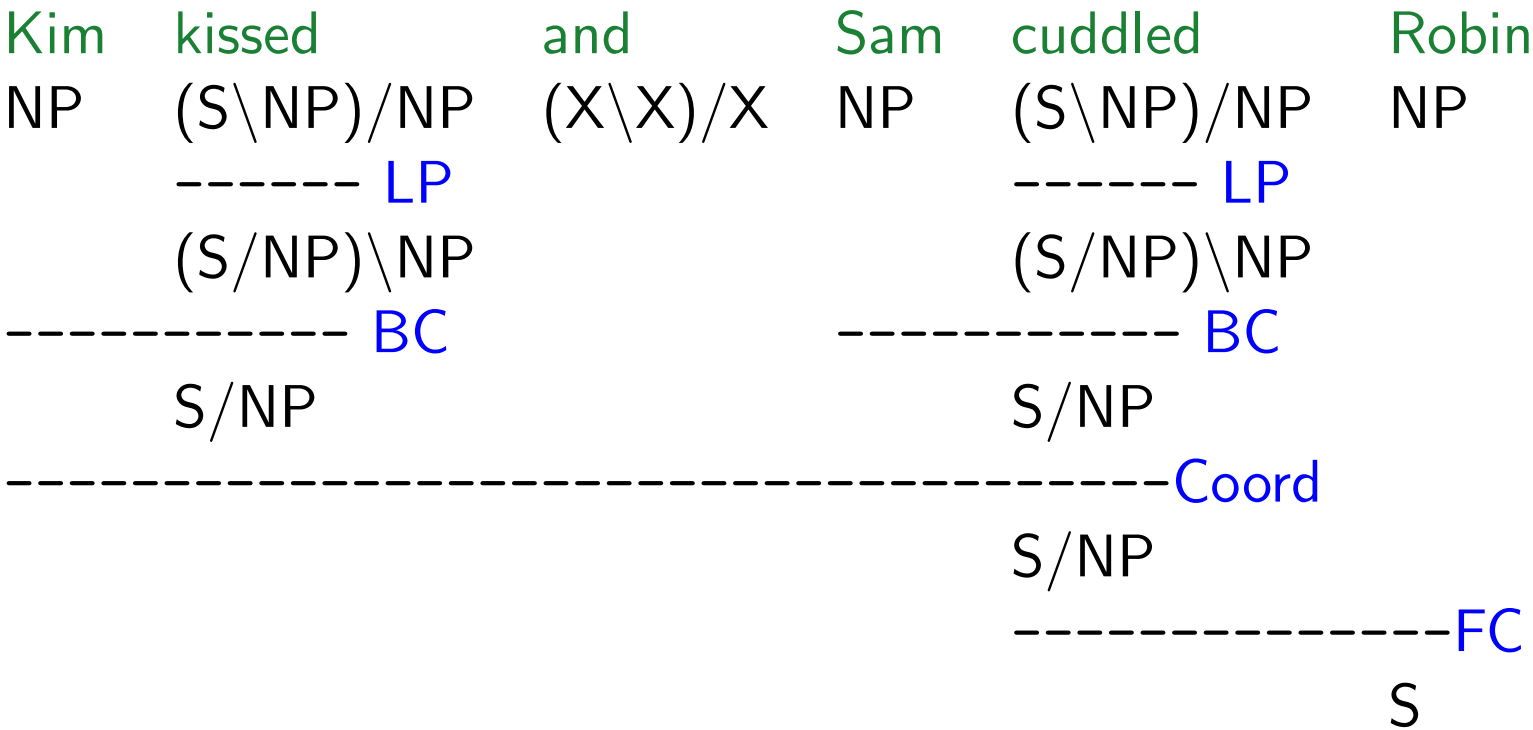
$(X|_1Y_1)\ldots|_nY_n \Rightarrow (X|_nY_n)|_1Y_1\ldots$

$\lambda\ y_n\ldots,y_1\ [X(y_1\ldots,y_n)] \Rightarrow \lambda\ \ldots y_1,y_n\ [X(y_1\ldots,y_n)]$

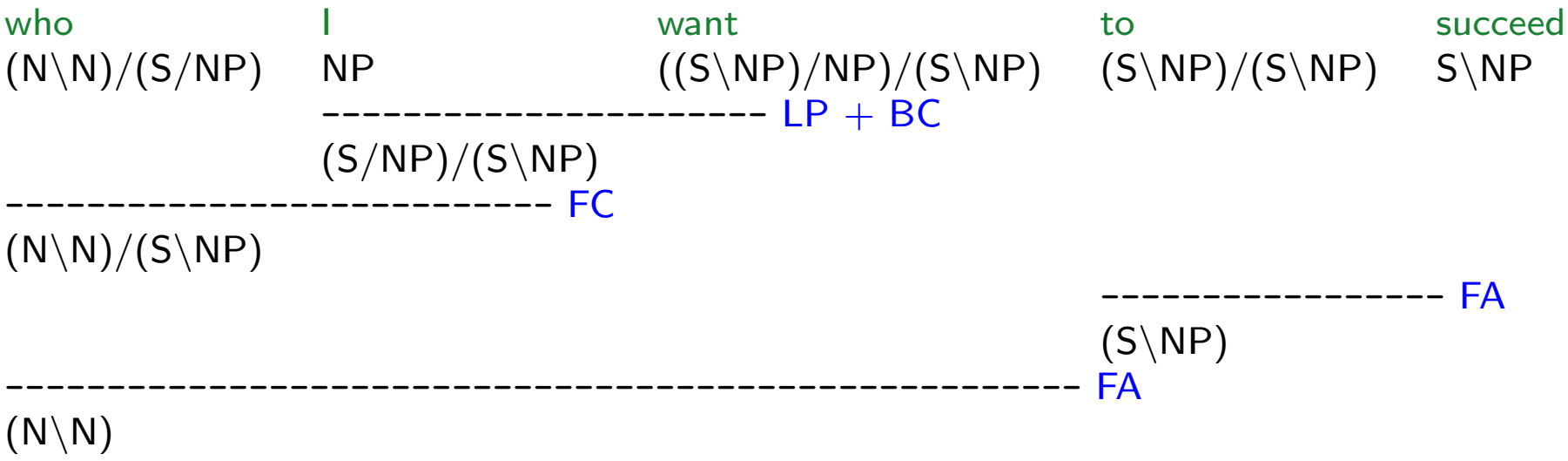
Derivation with Application

Kim	kissed	Sandy	in	Paris
NP	(S\NP)/NP	NP	((S\NP)\(S\ NP))/NP	NP
kim′	λ y,x kiss′(x y)	sandy′	λ y,P,x in′(y P (x))	paris′
	---------- FA		---------- FA	
	S\NP		(S\NP)\(S\ NP)	
	λ x kiss′(x sandy′)		λ P,x in′(paris′ P(x))	
	---------- BA			
	S\NP			
	λ x in′(paris′ kiss′(x sandy′))			
---------- BA				
S				
in′(paris′ kiss′(kim′ sandy′))				

Derivation with Permutation



Derivation with Composition



. . . who I want e to succeed

GCG Absolute (UG) Universals

- Compositionality, Productivity...
- Mild Context Sensitivity: nesting (a^n b^n, aabb), cross-serial (a^n b^n c^n, aabbcc), intersecting (a^n,b^n,c^n, cabbca)

- The guy Kim kissed smiled (A)
- Kim-NOM the house-DAT helped paint (A+C)
- document-ACC spy-DAT police-NOM journalist-NOM handed reported (A+C+P)
- The ...-ACC (<7) kissed / kissed the ...-ACC (>7) (S...\NP)/S

Genes, Memes, Signs

- Selection acts on the phenotype (organism)
- Inheritance of the genotype (genes)
- Selection acts on signs (constructions, words)
- Inheritance (Learning) of parameters
- Linguistic parameters categories, lexemes
- Well-defined (compared to memes)

Coevolution

- Symbiotic / Parasitic – predator/prey, cleaner fish
- Gene-Culture(Meme) – lactose tolerance / dairy farming
- Genetic Assimilation – Waddington, fruit flies, now lactose
- Gene-Language(Sign) – Baldwin Effect (Pinker & Bloom)
- Linguistic Ecological Niche – (Un)masking (Deacon)
- Language change – too fast for genes to keep up (Deacon)

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Language Agents

- Form-meaning pair: $fm_k = f_k + m_k$
- Language Agent: $(LAgt_i)$
  1. $< lg^j = LP(CCG - UG, fm_k)$
  2. $m_k = Parse(lg^j, f_k)$
  3. $f_k = Generate(lg^j, m_k)$
  4. $Age(0/1) >$
- Interactions, Communicative Success...
- Language = (En/Dec)Coding (Grammar) + Inference – cheap (Levinson), or expensive (Grice, Sperber/Wilson)?

Summary

- Modelling (E-)Language(s) as Dynamical / (Complex) (Adaptive) Systems
- Generalized Categorial Grammar as a Model of (I-)Language
- (E-)Language and the Language Faculty may have co-evolved
- Language Agent – UG(CCG) + LP + Gen + Pars (+ Inf.!!!)

Suggested Reading

Steedman, M. *The Syntactic Process* MIT Press, 2000.
Deacon, T. *The Symbolic Species: Coevolution of Language and the Brain*, Norton, 1997.
Johansson, S. *Origins of Language*, Benjamins, 2005
Briscoe, E.J. "Evolutionary Perspectives on Diachronic Syntax", In Pintzuk, S., Tsoulas, G. and Warner, A. *Diachronic Syntax: Models and Mechanisms*, OUP, 2000.
Briscoe, E.J. "Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device", *Language* 76.2, 2000.
www.cl.cam.ac.uk/users/ejb/