The Final-over-Final Constraint: An Evolutionary Linguistic Perspective

Ted Briscoe

Computer Laboratory
Natural Language and Information Processing Group
University of Cambridge

Cambridge Institute for Language Research Workshop
Jan 2010

An Analogy

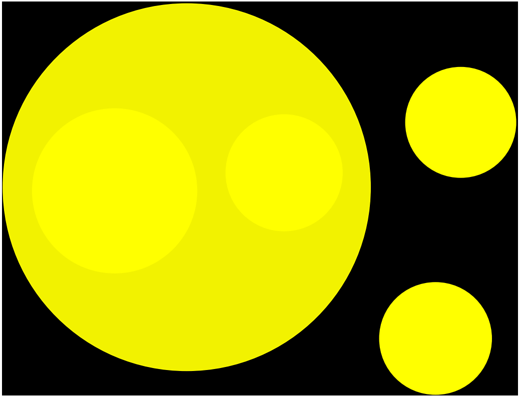
A drunk had lost his keys on the street and was frantically searching for them under a streetlamp. ‘Where did you drop them?’ asked a concerned passer by. ‘Over there’ he replied, indicating a spot 30 yards away. ‘So why are you looking here under the lamp?’ ‘The light is better here’.

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?

Universal Darwinism

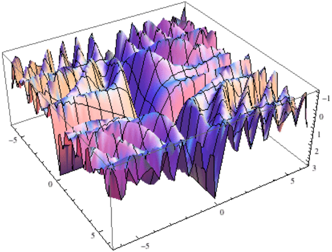
Languages don't just change they *evolve*. And children themselves are the rigged game. Languages are under powerful selection pressure to fit children's likely guesses, because children are the vehicle by which a language gets reproduced. Languages have to adapt to children's spontaneous assumptions...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

1. Linguistic Variation +
2. Language Acquisition +
3. Linguistic Selection =
4. Linguistic Evolution

Linguistic Selection

1. Learnability – frequency, interpretability, learning bias...
2. Expressiveness – processing economy, memorability, prestige...
3. Interpretability – processing efficiency, distance, ambiguity...

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



Generalized Categorial Grammar

Forward/Backward Application (F/B A):

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

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

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

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

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

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

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

```
Kim   kissed       and       Sam   cuddled       Robin
NP    (S\NP)/NP    (X\X)/X   NP    (S\NP)/NP     NP
      ------- LP                   ------- LP
      (S/NP)\NP                    (S/NP)\NP
----------- BC               ----------- BC
      S/NP                         S/NP
-------------------------------------Coord
                                   S/NP
                                   ---------------FC
                                                 S
```

Derivation with Composition

```
who            I              want                 to             succeed
(N\N)/(S/NP)   NP             ((S\NP)/NP)/(S\NP)   (S\NP)/(S\NP)  S\NP
               ---------------------- LP + BC
               (S/NP)/(S\NP)
---------------------------- FC
(N\N)/(S\NP)
                                                   ----------------- FA
                                                   (S\NP)
----------------------------------------------------- FA
(N\N)
```

. . . who I want e to succeed

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

Bayesian Parametric Learning of GCG

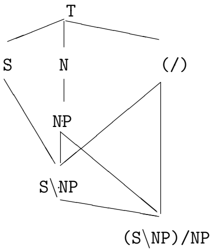
- Input – finite noisy form-meaning pairs (fm_n):
  Daddy gave you the sock throw′(daddy′ you′ x) ∧ sock′(x)
- Hypothesis Space – F/B A+C, L/D P + Cat. + Lex.
- Learning Bias / Occam's Razor – prior distribution on set of finite-valued parameters (A,C,P + Cat. set):
  $p(g \in G) = \prod_{param_i \in g} p(param_i = x)$
- Incremental Learning, posterior distribution given input:

  for $0 < i < n, argmax_{g \in G}\ p(g)\ p(fm_i \mid g)$

  $p(fm_i \mid g) = \prod_{param_j \in fm_i} p(param_j)$

  $p(param_j = x) = \frac{f(param_j = x) + 1}{f(param_j = X) + N}$

Parametric Specification of Category Sets



Finite Feature / Category Set:

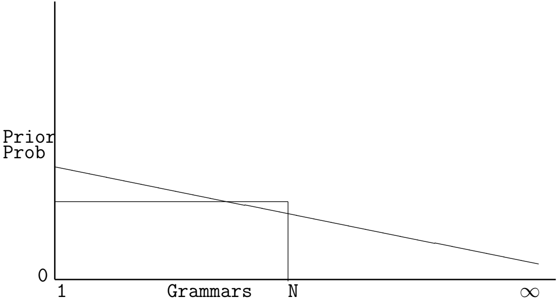
NP	=	[CAT=N, BAR=1, CASE=X, PERNUM=Y]
S	=	[CAT=V, BAR=0, PERNUM=X]
\NP	=	[DIR = left, CAT=N,...]
		$S_{pernum=x} \backslash NP_{pernum=x}$
		$S \backslash NP_{pernum=3sg} \sqcap NP_{case=nom} = NP_{3sg,nom}$

Chomskyan vs. Bayesian Learning

Learning Universal: Irregularity correlated with frequency

go+ed / went, ((S\IT)/NP)/S annoy, bother,...

Convergent Evolution: lrng biases walk thru' parameter space



(1,1)-Bounded Context Parser

Stack Cells		Lookahead	Input Buffer
2	1		
(who) (N\N)/(S/NP)	(you want) (S/NP)/(S\NP) S/(S\NP)	to (S\NP)/(S\NP)	succeed

Costs / cell	
4	2

3 Shifts, 1 Reduce to reach this configuration
Onset of the shift-reduce ambiguity at the first potential gap

Working Memory Cost Metric

After each parse step (Shift, Reduce, Halt):

1. Assign any new Stack entry in the top cell (introduced by Shift or Reduce) a cost of 1 multiplied by the number of CCG categories for the constituent represented (Recency/Recoding)
2. Increment every Stack cell's cost by 1 multiplied by the number of CCG categories for the constituent represented (Decay)
3. Push the sum of the current costs of each Stack cell onto the Cost-record (complexity at each step, sum = tot. Complexity)

Processing Complexity of Constructions / Sentences

- The students who the police who the reporters interviewed arrested laughed (161/547)
- The students who the reporters interviewed who the police arrested laughed (87)
- daB Peter dem Kunden den Kuhlschrank zu reparieren zu helfen versucht (294)
- daB Peter versucht dem Kunden den Kuhlschrank zu reparieren zu helfen (117)
- He donated the largest single sum ever given by a private individual to the university (C)
- He donated to the university the largest single sum ever given by a private individual (C+20)
- Short < Long (Dependencies & Constituents) – convergent evolution (heavy np shift, extraposition)

Tense-Verb-Object Cases

```
Aux              V                O              | O                Aux              V
(S|NP)/(S|NP)    (S|NP)/NP        NP             | NP               (S|NP)/(S|NP)    (S|NP)\NP
------------------- FC                           |                  ------------------- BC
                 (S|NP)/NP                       |                  (S|NP)\NP
                 --------------------FA          | ------------------BA
                                  S|NP           | S|NP
-------------------------------------------------+-------------------------------------------------
V                Aux              O              | O                V                Aux
(S|NP)/NP        (S|NP)\(S|NP)    NP             | NP               (S|NP)\NP        (S|NP)\(S|NP)
--------------------- BC                         | ------------------- BA
                 (S|NP)/NP                       |                  S|NP
                 -----------------------FA       |                  ------------------BA
                                  S|NP           |                  S|NP
-------------------------------------------------+-------------------------------------------------
*V               O                Aux            | Aux              O                V
(S|NP)/NP        NP               (S|NP)\(S|NP   | (S|NP)/(S|NP)    NP               (S|NP)\NP
-------------------- FA                          |                  ---------------------- BA
                 S|NP                            |                  S|NP
                 ------------------------BA      | ---------------------FA
                                  S|NP           | S|NP
```

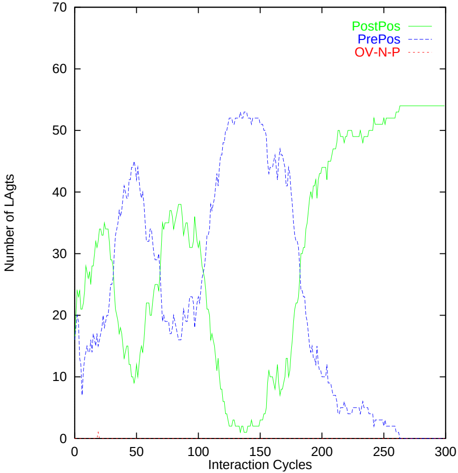
LP - Complexity

- Hierarchy:
  OVT < TVO (Comp.) < OTV (Less-Incr.) < VTO (Non-Harm.) < *VOT (O-Non-Incr.) < TOV (Non-Incr.)
- Extraposition:
  *VOT → VTO but TOV → TVO
- Historical Pathways:
  Down Hierarchy < more probable: e.g.
  OVT → ?TOV ⇒ TVO
  OVT → *VOT ⇒ TVO
  Tense less stable than Verb:
  OTV ⇒ OVT
  VTO ⇒ TVO

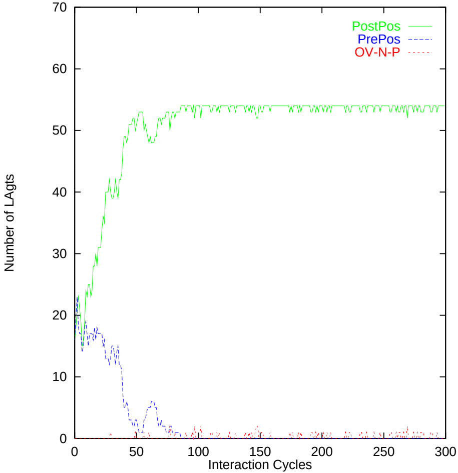
UG - Constraint

- Feature-based FoFC Constraint:
  *((Head$_\alpha$ Obj) Head$_\alpha$)
  *((X/Y Y) X$'\backslash$X)
- OBJDIR:
  X[OBJDIR right]/Y[OBJDIR X]
  X$'\backslash$X[OBJDIR left]])
- Non-local Feature:
  *((...(Head$_\alpha$ Obj)) Head$_\alpha$)
  Like Gap features in GPSG/HPSG
- Increased overall expressive power despite enforcing FoFC
- Black Swans – 'absence of evidence is not evidence of absence' in (a sample of) attested languages

OV+Prep/Post without processing costs



OV+Prep/Post with processing costs



Conclusions and References

- FoFC is hard to formalise as a constraint within UG without increasing generative capacity and thus learning complexity
- FoFC violation is predicted to be dispreferred because it is both disharmonic and non-incremental and is not ameliorated by extraposition
- Convergent evolution of languages is an alternative non-UG / non-nativist explanation for (apparently) exceptionless universals

Steedman, M. *The Syntactic Process* MIT Press, 2000.
Briscoe, E.J. "Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device", *Language* 76.2, 2000.
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. and Buttery, P. "Linguistic Adaptations for Resolving Ambiguity", in

Procs. of Evolang 7, World Scientific, 2008 (eds.) Smith, Smith & Ferrer i Cancho

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