# Second Language Acquisition and the Final-over-Final Constraint

Ted Briscoe

Computer Laboratory
Natural Language and Information Processing Group
University of Cambridge

ENS, Paris
Mar 2014

# Modelling Creolisation

- Pidgin Lg – randomly ordered {S,O,V} (substratum influences), no NP internal grammar, no embedding
- Biased Prior Learners – learner population reliably fixates on SVO subset lg after 1 generation
- Superstratum Lg – some more complex categories from SVO ('English') or SOVv2 ('Dutch') leads to full SVO lg ('Hawaian Creole / Saramaccan') after 2 generations
- Prior bias and input distribution ('linguistic demographics') predict creolisation (15–28% learners, 10% SVO/SOVv2 adults, 75–62% pidgin SVO adults)
- How does the pidgin emerge? – Second Language Acquisition

# Modelling Creolisation

- Pidgin Lg – randomly ordered {S,O,V} (substratum influences), no NP internal grammar, no embedding
- Biased Prior Learners – learner population reliably fixates on SVO subset lg after 1 generation
- Superstratum Lg – some more complex categories from SVO ('English') or SOVv2 ('Dutch') leads to full SVO lg ('Hawaian Creole / Saramaccan') after 2 generations
- Prior bias and input distribution ('linguistic demographics') predict creolisation (15–28% learners, 10% SVO/SOVv2 adults, 75–62% pidgin SVO adults)
- How does the pidgin emerge? – Second Language Acquisition

# Modelling Creolisation

- Pidgin Lg – randomly ordered {S,O,V} (substratum influences), no NP internal grammar, no embedding
- Biased Prior Learners – learner population reliably fixates on SVO subset lg after 1 generation
- Superstratum Lg – some more complex categories from SVO ('English') or SOVv2 ('Dutch') leads to full SVO lg ('Hawaian Creole / Saramaccan') after 2 generations
- Prior bias and input distribution ('linguistic demographics') predict creolisation (15–28% learners, 10% SVO/SOVv2 adults, 75–62% pidgin SVO adults)
- How does the pidgin emerge? – Second Language Acquisition

# SLA and L1 Transfer / Basic Variety

- Word Order: Turkish L1 (SOV) – I something eating, Finish no (Corrected rapidly)
- Dative: French L1 – *I gave Kim a dog, English L1 J'ai donné Kim un chien
- Function words: Turkish L1 (bir 'one') – So brain is already shaped
- Morphology: Turkish L1 (synthetic) – I scan some in the computer

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f' m'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Extensions to Learning Procedure (LP)

1. L1/L2 Perceptability: polysyllabic lexemes > monosyllabic stressed lexemes > free morphemes > clitics > bound morphemes > inflections (opposite of production economy, functioning analogously to memory cost in acquis. model)
2. L2 Input:
   - Form-meaning pair: $fm_k = f_k + m_k$
   - Translation pair: $fm_k = f'm'_{k'}$ (L2=L1)
   - L2 Input pair: $f_k + m'_{k'}$ (L2f=L1m)
3. L2 Starting Point: L1 parameter settings as 'defaults' (i.e. reset mature parameter estimates to minimally-biased least-confident settings and learn from evidence
4. L1/L2 Communicative success more important than grammatical fidelity – the perceptability hierarchy is also a hierarchy of semantic informativeness

# Old / Middle English Change

Repeated Migrations:
the northern dialect of English most likely became a CP-V2 language under the extensive contact it had with medieval Scandinavian... The linguistic effect of this combination of population movement and population mixture was extensive, comparable in some ways to the pidginization/creolization phenomena of more recent centuries, though not as extreme... imperfect second language learners... were a sufficiently large fraction of the population... to pass on their mixed language to succeeding generations (Kroch & Taylor, 1997:318f)

Loss of infl. morphology, loss of (verb) movement
= morph./syn. trade-off or param. (re)setting?

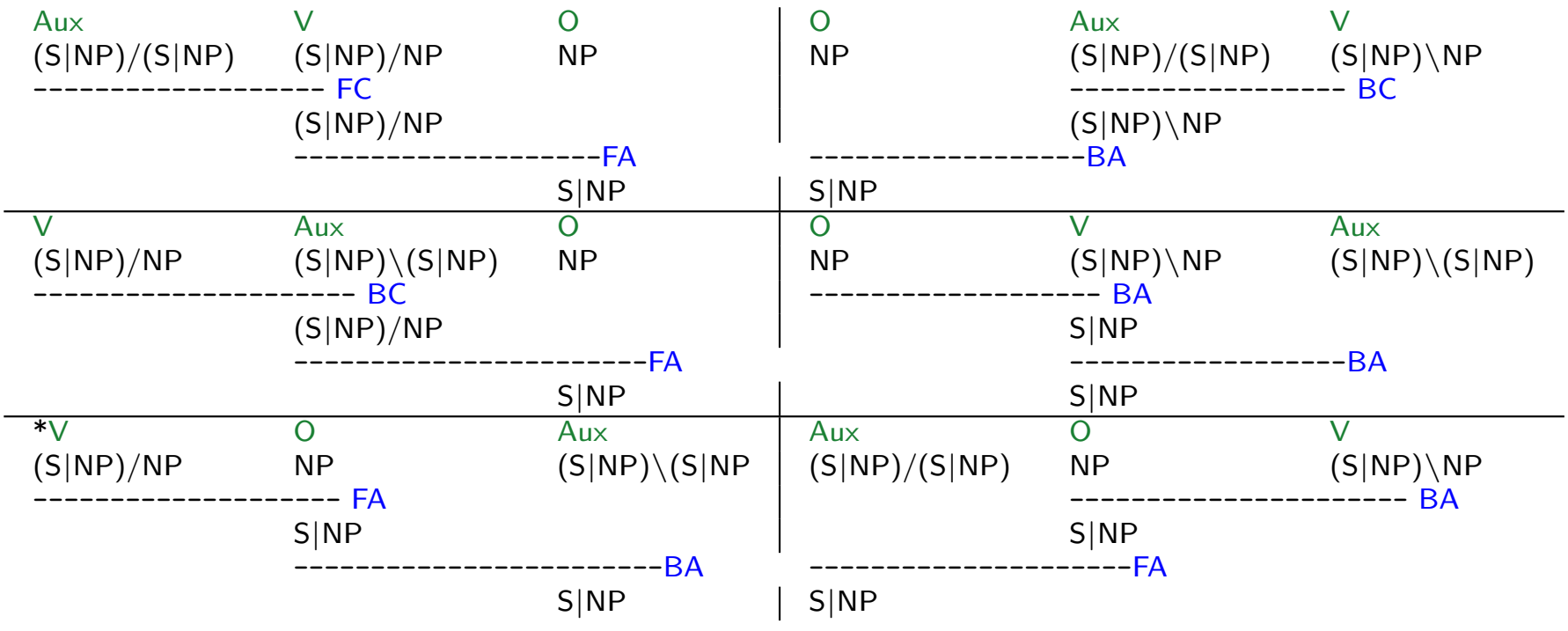
# Proportion of L2 Learners and Case Erosion (Bentz & Winter)

- WALS Database, 226 lgs from diverse lg families with L2 learner info
- Proportion of L2 speakers inversely correlates with number of cases (regardless of whether L1 contact lg has case)
- L2 speakers incorporated into Roman Empire – Latin $\rightsquigarrow$ Vulgar Latin $\rightsquigarrow$ Romance (fixed word order)

# The Final-over-Final Constraint

- A head-final phrase cannot dominate a head-initial phrase of same type (Biberauer et al)
- A weakening of head harmony principle to retain an absolute universal (UG)
- Rules out e.g. ((Vb Obj) Aux), ((Vb Obj) Comp) – Aux, Comp, Vb all V+

# Auxiliary-Verb-Object Cases



# LP(UG) + WMC - Complexity Predictions

- Hierarchy:
  OVA $<$ AVO (Comp.) $<$ OAV (Less-Incr.) $<$ VAO (Non-Harm.) $<$ *VOA (O-Non-Incr.) $<$ AOV (Non-Incr.)
- Extraposition (Long-last):
  *VOA $\rightarrow$ VAO but AOV $\rightarrow$ AVO
- Historical Pathways:
  Down Hierarchy $<$ more probable: e.g.
  OVA $\rightarrow$ ?AOV $\Rightarrow$ AVO
  OVA $\rightarrow$ *VOA $\Rightarrow$ AVO
  Tense Auxiliaries less stable than Verb:
  OAV $\Rightarrow$ OVA
  VAO $\Rightarrow$ AVO

# Formalising as UG Constraint

- Feature-based FoFC Constraint:
  *((Head$_\alpha$ Obj) Head$_\alpha$)
  *((X/Y Y) X′\X)
- OBJDIR:
  X[OBJDIR right]/Y[OBJDIR X]
  X′\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

# Typological Predictions

- Predictions for ordering of A,V,O are similar to those for other typological non-harmonic universals
- Mixed (non-harmonic) heads of all category types dispreferred
- Ordering long before short or having long intervene between short dispreferred
- *((...(Head$_\alpha$ XP)) Head$_\beta$) and (...Head$_\beta$ (XP Head$_\alpha$)) are equally uncommon but not completely unattested (0–5%)

# Artificial Language Learning Experiments

- Adj-Noun-Num ordering: *((A N) Nu), ?(Nu (N A)) vs. ((Nu (A N)), ((N A) Nu)
- Culbertson et al – mixed more difficult and first even more difficult to learn (explain in terms of Bayesian priors favouring harmony, regularisation, and 'substantive' learning biases: A-N $\rightarrow$ Nu-N)
- Goldberg – the further bias against the FoFC-violating (Adj=Head) word order is not due to a substantive bias but to a L1 transfer effect (subjects spoke English or Spanish)

# Conclusions

- Instability / Change predicted when a sign is complex to learn, perceive or process
- Language (change) is epiphenomenal grounded in (changing) interactions between language users – demography
- (Un)Folding of Language(s) via 'grammaticalisation' in homogeneous communities and 'creolisation' in hetrogeneous communities?
- L1/L2 acquisition same process, different context? – model it!
- 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
- Convergent evolution given learning / processing selection pressures on languages is a better non-UG explanation for (nearly!) exceptionless universals

# Readings

Biberauer, T., Newton, G., and Sheehan, M. "Impossible changes and impossible borrowings: the Final-over-Final Constraint" in *Continuity and Change in Grammar* (eds.) Breitbarth, A. et al, Benjamins, 2010

Goldberg, E. "Substantive learning bias or an effect of familiarity? Comment on Culbertson et al." *Cognition* 127, 2012

Bentz, C. and Winter, B. "Languages with more 2nd language learners tend to have smaller case systems" *Language Dynamics and Change* 2013

Briscoe, E.J. "Grammatical Acquisition and Linguistic Selection", in *Linguistic evolution through language acquisition: formal and computational models*, (ed.) Briscoe, E.J., Cambridge University Press, pp255-300, 2002

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