Rate-dependent dynamics of grammatical diffusion on social networks

Robert Daland & Janet B. Pierrehumbert

Presented to the Northwestern/EHESS workshop
Dynamics and complexity in people and societies

Oct. 23, 2007
"Hi' mother a chimpanzee"
Grammatical variation

- Varies within and across speakers
  - Hi' mother (is) a chimpanzee. (null copula)
  - John said (that) Mary kisses better. (optional-
    that)
- Present in all linguistic domains
  - Coker ~ coca (phonology)
  - singing ~ singin' (morphophonology)
- Constrained by understandability
  - John better Mary kisses that said. (head-final)
  - John said that Mary kisses better. (head-initial)
The logical problem of coordinating language systems (grammars)

• How do social and cognitive processes interact to coordinate grammars within a group?
  − Learners are exquisitely sensitive to the input distribution
    • or they could not coordinate their grammars with their neighbors
  − Learners do not faithfully reproduce the input distribution
    • or else dialects and languages would never diverge over the course of history
Plan

• Social network model
• Three cases
  – Linear diffusion
  – Weakly nonlinear diffusion
  – Strongly nonlinear diffusion
• Conclusions
Plan

• Social network model
  • Three cases
    – Linear diffusion
    – Weakly nonlinear diffusion
    – Strongly nonlinear diffusion
  • Conclusions

• Preferential attachment
• Production and learning rules
• Approximations
• Representation as a complex dynamical system
Preferential attachment

\[ \langle k \rangle = 2 \]

- Add nodes with \( \langle k \rangle \) new bidirectional links, until \( n \) nodes total
- Higher probability of linking to existing node with more links
Production and learning rules

• Probabilistic grammar, discrete utterance
  − discrete utterance types $\sigma \in \{0, 1\}$
  − produced according to observed probability $g$
  − cognitive biases $f$ affect output distribution
  − $p(\sigma = 1) = f(g)$

• Grammar update
  − idea: on hearing $\sigma$, update $g$ to increase likelihood of $\sigma$
  − linear reward/penalty
  − $g(t+1) = \alpha \cdot \sigma + (1-\alpha) \cdot g(t)$
Example nonlinearity

- **Bistability constraint**
  - both homogeneous categorical solutions are stable
  - $\langle g \rangle^0 \in \{0, 1\} \Rightarrow \langle g \rangle^\infty = \langle g \rangle^0$

- **When $g = .2$**
  - faithful: $p(1) = .2$
  - weak: $p(1) = .12$
  - strong: $p(1) = .02$

- **Theory**
  - Cucker, Smale, & Zhou, 2004; Niyogi, 2006

- **Simulation**
  - Crawford, 2007; Pearl & Weinberg, 2007; Troutman, Goldrick, & Clark, 2007
Approximations

• If the learning rate is small
  – No single example has a large effect
  – So the exact order of inputs is irrelevant
• Assume simultaneous updating:
  \[ g^{t+1} \approx (\alpha A) \cdot \sigma^t + (1-\alpha K) \cdot g^t \]
• Assume mean-field output:
  \[ \sigma^t \approx \langle \sigma^t \rangle = f(g^t) \]
• Put it all together, and...
Voila!

\[ g^{t+1} = \alpha A \cdot f(g^t) + (I - \alpha K) \cdot g^t \]
Plan

- Social network model
- Three cases
  - Linear diffusion
  - Weakly nonlinear diffusion
  - Strongly nonlinear diffusion
- Conclusions

- Symmetric pseudo-Markov chain
- Convergence properties
- Effects of
  - social structure
  - size and connection density
Symmetric pseudo-Markov chain

- Production is faithful to perceived input distribution
  - \( f(g) = g \)
- Substitute and simplifying:
  - \( g^{t+1} = \alpha A \cdot g^t + (I - \alpha K) \cdot g^t = M \cdot g^t \)

- \( M \) is symmetric and row-stochastic
- (always assume \( A \) is irreducible so \( M \) is too)
Convergence properties

• Symmetric row-stochasticity of M guarantees
  – principal eigenvalue is 1 and principal eigenvector points in direction (1, 1, ..., 1)
  – $g^\infty = \lim_{t \to \infty} g^t = g^\infty \cdot (1, 1, ..., 1), \quad g^\infty = \langle g^0_i \rangle$

• All agents converge to distribution $g^\infty$, which is the average of each agent's initial $g$

• Rate of convergence depends on social structure
  – $\tau_{\text{coupling}} = -1/(\lg |\lambda_2|)$ (independent of $g^0$)
Social structure

Distribution of convergence times

Gross structural properties fixed:
\[ n = 40 \]
\[ \langle k \rangle \approx 4 \]
Size and connection density

Effect of $n$ and $<k>$ on convergence rate

Mean number of neighbors

Number of agents

TGC
Discussion

• Main result
  – Diffusion is rapid and robust to changes in network size or other parameters (except at small k)

• Connection to epidemiology
  – Disease propagation is faster in preferential attachment network (Anderson & May, 1991)

• Sociolinguistic implications
  – Language change faster in densely connected areas
  – Compatible with claims in the literature
  – Iffiness: What is a connection?
Plan

- Social network model
- Three cases
  - Linear diffusion
  - Weakly nonlinear diffusion
  - Strongly nonlinear diffusion
- Conclusions

- Mean-field approximation
- Convergence properties
- Calculating $\tau$ cognitive
Mean-field grammar approximation

- **Problem:** multi-variable nonlinear system hard to analyze
- **Solution:** assume mean agent is representative
  \[ \langle g \rangle^{t+1} = \alpha \langle k \rangle \cdot f(\langle g \rangle^t) + (1-\alpha \langle k \rangle) \cdot \langle g \rangle^t \]
  \[ = F(\langle g \rangle^t) \]
- reduces to single-dimensional system
- only valid when \( \tau_{\text{coupling}} \ll \tau_{\text{cognitive}} \)
Convergence properties

• Bistability + nonlinearity of $f$ implies a repellor $b$
  - $\exists b \in (0,1) \ni f(b) = b$ and $f'(b) > 1$

• If $b$ is unique:
  - $\langle g \rangle^\infty = 0$ whenever $\langle g \rangle^0 < b$:
  - $\langle g \rangle^\infty = 1$ whenever $\langle g \rangle^0 > b$
Calculating $\tau_{\text{cognitive}}$

- **Lyapunov exponent**
  \[ \lg \lambda_T(F, \langle g \rangle^0) = \langle \lg F'(\langle g \rangle^t) \rangle \quad \text{where} \quad T = \begin{bmatrix} \tau \\ \tau_{\text{coupling}} \end{bmatrix} \]

- **Time constant defined analogously**
  \[ \tau_{\text{cognitive}} = -1 / \lg \lambda_T(F, \langle g \rangle^0) \]

- **May be possible to estimate F empirically with artificial language studies** *(a la Hudson Kam & Newport, 2005)*
Plan

- Social network model
- Three cases
  - Linear diffusion
  - Weakly nonlinear diffusion
  - Strongly nonlinear diffusion
- Conclusions

- Necessary conditions for dialect formation
- A novel prediction
Necessary conditions

• Dialect splitting
  – cannot occur with no cognitive pressure
  – cannot occur when $\tau_{\text{coupling}} << \tau_{\text{cognitive}}$
  – thus can only occur when $\tau_{\text{cognitive}} \lessapprox \tau_{\text{coupling}}$

• Sensitive to initial conditions
  – regions must lie on opposite sides of bifurcation
A novel prediction

• Stable variation only predicted when cognitive pressure weak or non-existent

• Cognitive pressure won't change when dialects split on an independent linguistic dimension

• Linguistic variables in stable variation will have a similar distribution across different dialects
  - e.g. optional-\textit{that}
Plan

• Social network model
• Three cases
  – Linear diffusion
  – Weakly nonlinear diffusion
  – Strongly nonlinear diffusion
• Conclusions
Conclusions

- Social dynamics of language transmission robustly coordinate grammars across a network.

- Cognitive biases such as memory limitations result in a preference for categorical grammars.

- Dialect splitting should only occur when time course of cognitive effects is fast relative to time course of social coordination process.
Thank you!

- Celina Troutman, Matt Goldrick, and Brady Clark
- Lisa Pearl
References


