Typological consequences of agent interaction

Coral Hughto Robert Staubs Joe Pater

University of Massachusetts Amherst

NECPhon 8
November 15, 2014

In standard generative grammar: Grammatical theories are constructed to generate all and only possible languages.

In standard generative grammar: Grammatical theories are constructed to generate all and only possible languages.

Some systems are permitted by the theory, others are not. No distinction is made within either class.

Standard goal of learning theories: Show how the systems generated might be learned.

Standard goal of learning theories: Show how the systems generated might be learned.

No independent role of learning in typological modeling.

We can do better than this: Explain relative frequency based on relative learnability—combining a learning theory with a grammatical theory (e.g. Heinz 2009, Pater and Moreton 2012, Staubs 2014).

We can do better than this: Explain relative frequency based on relative learnability—combining a learning theory with a grammatical theory (e.g. Heinz 2009, Pater and Moreton 2012, Staubs 2014).

Individual learners acquire particular patterns faster or slower based on how learning and grammar interact.

Today we'll focus on a third model bias.

Today we'll focus on a third model bias.

This bias emerges from interaction between agents, both within and across generations.

Today we'll focus on a third model bias.

This bias emerges from interaction between agents, both within and across generations.

We show consequences particularly for probabilistic models of grammar such as Maximum Entropy (Goldwater and Johnson 2003).

Our focus today: Interaction and transmission tend to reduce variability.

Our focus today: Interaction and transmission tend to reduce variability.

This happens in two fundamentally different network assumptions: iterated and interactive learning.

We show that these models show emergent tendencies towards:

1. Categorical outcomes
2. Lexical contrast
3. Avoidance of cumulativity

Error-driven learning

MaxEnt SGA (perceptron, HG-GLA; Jager 2007, Boersma and Pater 2014):

Error-driven learning

MaxEnt SGA (perceptron, HG-GLA; Jager 2007, Boersma and Pater 2014):

New Weights $=$
Old Weights $+ \eta \times$ (Learner Violations $-$ Teacher Violations)

Where η is some assumed learning rate.

Iterated Learning

Iterated learning models present a simplified model of language change

These models are based on the observation that language change happens over time: children's grammars are not exactly the same as their parents'

Agents in this model are arranged in a chain with one learner per "generation"

$$L_1 \rightarrow L_2 \rightarrow \ldots \rightarrow L_n$$

Each agent in a chain learns its language from the previous generation and then teaches it to the next (Kirby and Hurford 2002, Griffiths and Kalish 2007)

Typologically common languages coincide with languages which are stable (transmitted faithfully) under this learning model

Agents in an iterated learning chain preserve categorical grammar states better/longer than more variable grammars

This trend towards categoricity emerges through the transmission of languages between agents, without needing to encode a bias for categoricity within each agent

Interactive Learning

Interactive learning models present a simplified model of language generation (Dediu 2009, Pater and Moreton 2012)

A number of agents interact with and learn from each other:

$$L_1 \leftrightarrow L_2$$

From these interactions, a shared grammar emerges

This model is based on the observation that language change is a social phenomenon

An individual's language use continues to change over time, and their language use is affected by that of their social network

Probabilistic typological trends are reflected in the rate at which the agents generate particular systems under this model

The shared grammars developed by agents in an interactive learning model tend to be categorical

These effects are emergent properties of the model, and don't require any specifically encoded learning biases

Iterated or Interactive?

Iterated learning models emphasize the importance of the effect of transmission of language between generations (from adults to children), setting aside the social, interactive aspect of language learning

Interactive learning models emphasize the influence of peers on language development, setting aside the influence from adult language users

Both of these models are overly simplified; human language learning is probably influenced by both types of interaction

Categoricity

An interactive learner, starting with equal probabilities on candidates, will tend toward weights giving categorical outcomes.

Categoricity

An interactive learner, starting with equal probabilities on candidates, will tend toward weights giving categorical outcomes.

Categoricity tableau

	*A	*B
A	-1	
B		-1

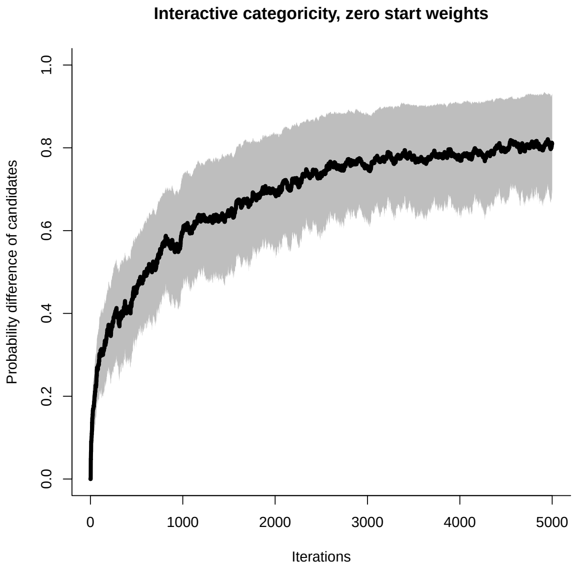
Categoricity

An interactive learner, starting with equal probabilities on candidates, will tend toward weights giving categorical outcomes.

Categoricity tableau

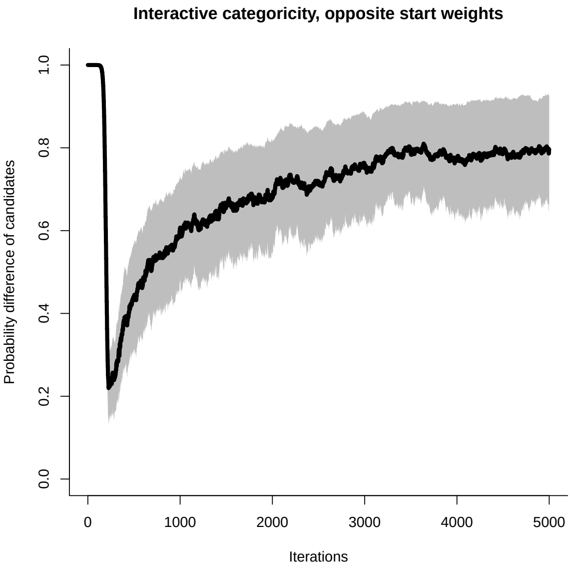
	*A	*B
A	-1	
B		-1

Dark lines with gray: means of 100 runs with standard deviations. Learning rate 0.1.


Interactive categoricity, zero start weights
Probability difference of candidates
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Iterations

The starting distribution is not crucial.

The learners converge on a shared categorical outcome even if they initially categorically disagree.


Interactive categoricity, opposite start weights
Probability difference of candidates
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Iterations

Interactive learning with probabilistic grammars, starting with non-categorical grammars:

Interactive learning with probabilistic grammars, starting with non-categorical grammars:

1. Errors can push the agents towards either more or less categorical states.

Interactive learning with probabilistic grammars, starting with non-categorical grammars:

1. Errors can push the agents towards either more or less categorical states.
2. As the agents drift into categorical grammars, they change less and less.

Interactive learning with probabilistic grammars, starting with non-categorical grammars:

1. Errors can push the agents towards either more or less categorical states.
2. As the agents drift into categorical grammars, they change less and less.
3. The system spends most of its time in categorical states.

Interactive learning with constraints (e.g. MaxEnt, Noisy HG; Boersma and Pater 2014):

1. Errors can push the agents towards either more or less categorical states.
2. As the agents drift into categorical grammars
   1. They change less and less.

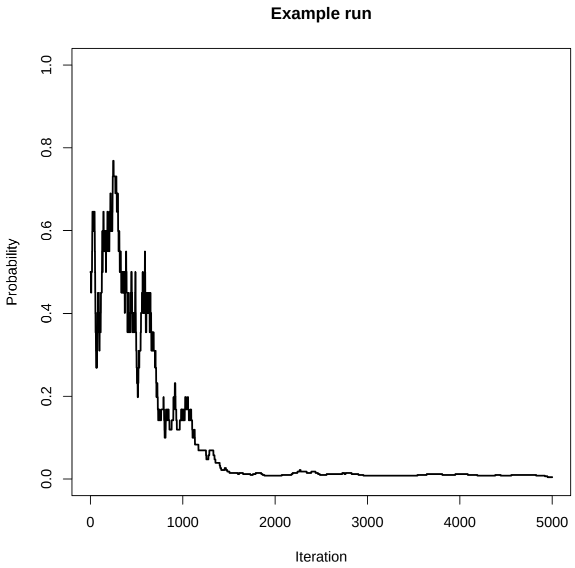
Interactive learning with constraints (e.g. MaxEnt, Noisy HG; Boersma and Pater 2014):

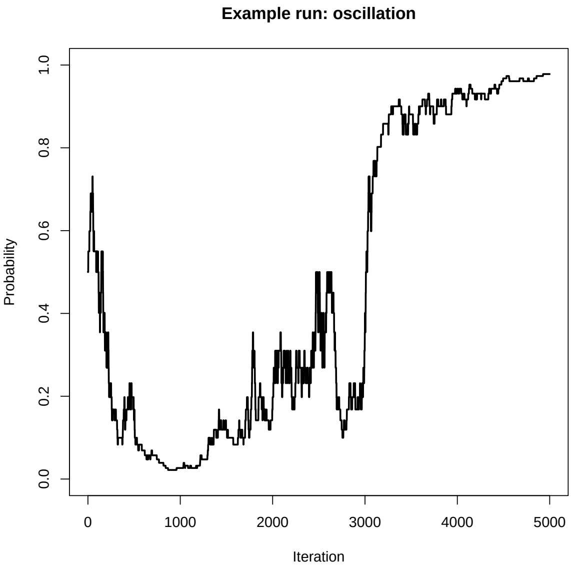
1. Errors can push the agents towards either more or less categorical states.
2. As the agents drift into categorical grammars
    1. They change less and less.
    2. The effective change to probability from a change in weights shrinks.

Interactive learning with constraints (e.g. MaxEnt, Noisy HG; Boersma and Pater 2014):

1. Errors can push the agents towards either more or less categorical states.
2. As the agents drift into categorical grammars
    1. They change less and less.
    2. The effective change to probability from a change in weights shrinks.
3. The system spends most of its time in categorical states.

(cf. Wedel 2007 on models where a positive feedback loop creates similar pressures)


Example run
Probability
Iteration
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000


Example run: oscillation
Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Iteration

Part of this pressure (at least) is present in iterated learning as well (see e.g. Dediu 2009, p. 555).

Part of this pressure (at least) is present in iterated learning as well (see e.g. Dediu 2009, p. 555).

Thus iterated learning can show a pressure for increasing categoricity over generations.

Part of this pressure (at least) is present in iterated learning as well (see e.g. Dediu 2009, p. 555).

Thus iterated learning can show a pressure for increasing categoricity over generations.

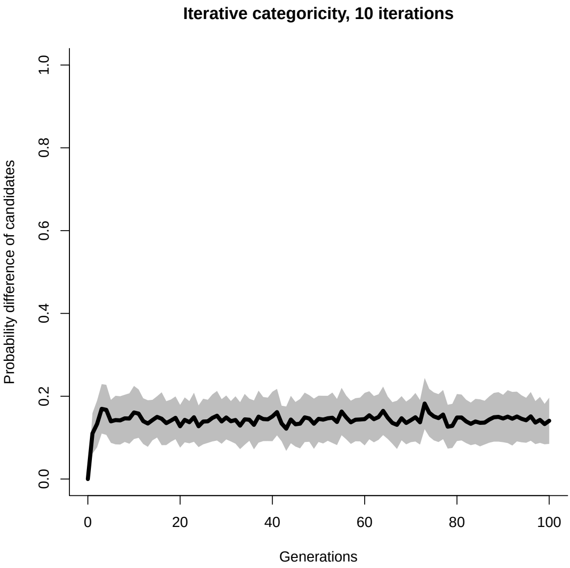
This requires enough learning to happen in each step in order to maintain the "emerged" categoricity.

Terminology reminder

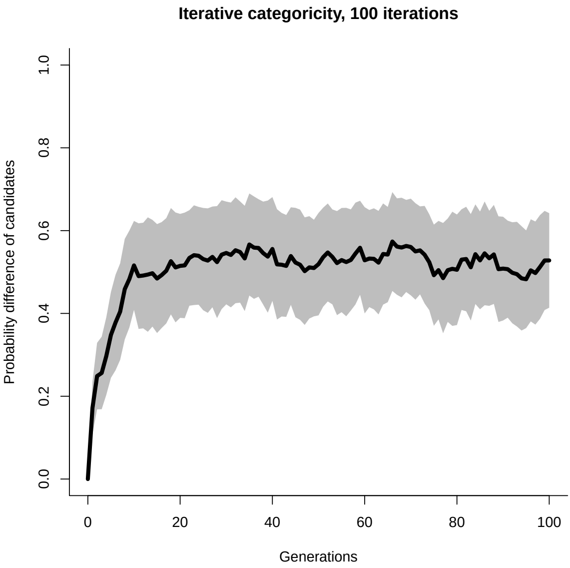
1. iteration: when a datum is exchanged between two agents
2. generation: when a new agent learns from another for a number of iterations

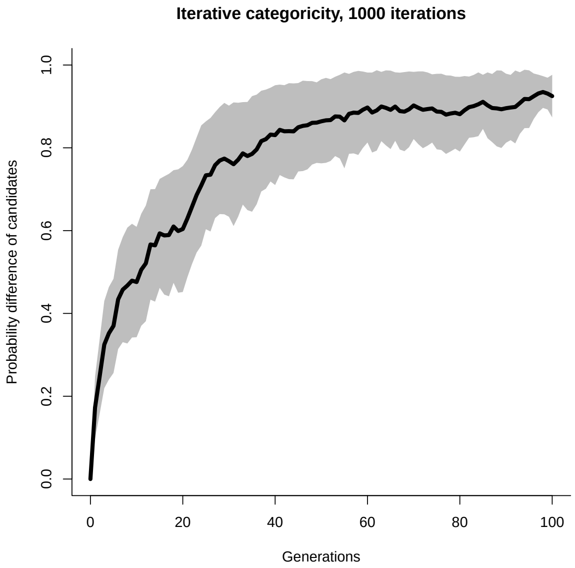
Thus iterations are relevant to both iterative and interactive.

Generations are not clearly important to interactive learning.



Iterative categoricity, 10 iterations
Probability difference of candidates
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Generations


Iterative categoricity, 100 iterations
Probability difference of candidates
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Generations


Iterative categoricity, 1000 iterations
Probability difference of candidates
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Generations

Cumulativity

In a weighted-constraint grammar, constraint violations are cumulative

The optimal candidate is the one whose Harmony score is closest to zero, but the particular combination of constraint weights and violations doesn't matter

- A candidate which incurs one violation of a constraint with a weight of 6 has the same Harmony score as a candidate which incurs two violations of a constraint with a weight of 3 $(1*6 = 2*3 = 6)$

Constraint cumulativity has been cited as a problem for weighted-constraint grammars, as it makes undesirable typological predictions (e.g. Legendre et al. 2006).

If cumulativity effects exist, it seems they might be uncommon.

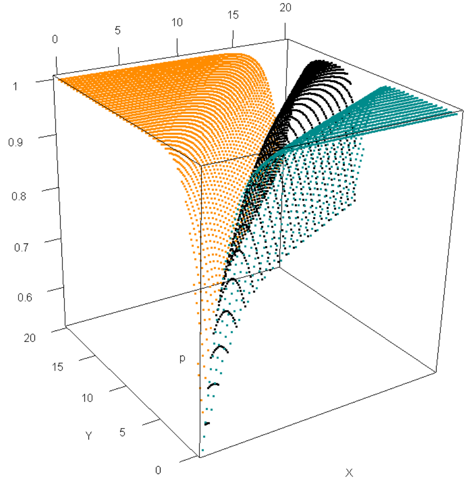
If either fact is true, we should worry about a model that treats cumulative languages identically with non-cumulative ones.

Cumulativity tableaux

	3	2	
	X	Y	H
→A		-1	-2
B	-1		-3
→C	-1		-4
D		-2	-6

In an interactive learning model, the agents strongly tend away from cumulative patterns

One reason: many cumulative weightings are intermediate and non-categorical (Carroll 2012).



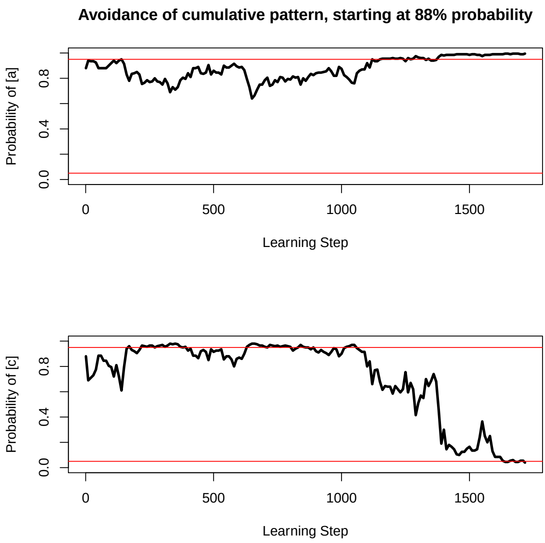
Cyan and **orange**: no cumulativity effect. **Black**: cumulativity.

Cumulativity

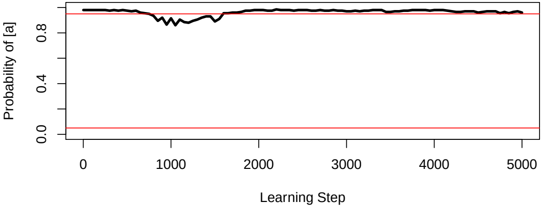
A simulation with two agents beginning with constraint weights at zero, run 1000 times, produced no cumulative patterns

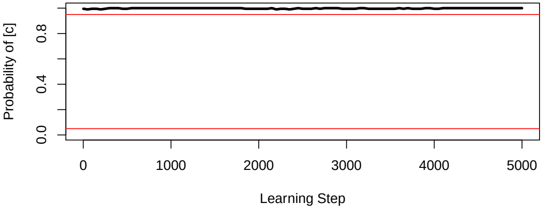
	X	Y
A		-1
B	-1	
C	-1	
D		-2

Language	Count
A, D	312
B, C	688
A, C	**0**


Avoidance of cumulative pattern, starting at 88% probability
Probability of [a]
0.0
0.4
0.8
0
500
1000
1500
Learning Step
Probability of [c]
0.0
0.4
0.8
0
500
1000
1500
Learning Step

Maintaining cumulative pattern, starting at very high probabilities





Cumulativity

Carroll (2012) analyses the real-world typology of contrasts between /s/ and /ʃ/, finding the following distribution of languages:

Contrast Type	Proportion
Total Neutralization	44.0%
Full Contrast	37.0%
Complementary Distribution	10.3%
Contextual Neutralization	8.2%
"Elsewhere" Neutralization	0.5%

Cumulativity

The "Elsewhere" Neutralization pattern is representable as a cumulative pattern in a weighted-constraint grammar, and is largely underrepresented in the typology

Carroll (2012) attempts to account for this skew away from cumulative patterns through encoding various biases into a MaxEnt learner, but doesn't find a solution that fits the data as well as desired

The interactive learning model presented here derives the avoidance of cumulative patterns that Carroll was looking for, without needing to encode specific learning biases

Contrast

The pressure for categoricity can be extended into a pressure for contrast.

Contrast

The pressure for categoricity can be extended into a pressure for contrast.

Now we have agents pronouncing different meanings, not just uninterpretable strings.

Contrast

The pressure for categoricity can be extended into a pressure for contrast.

Now we have agents pronouncing different meanings, not just uninterpretable strings.

We add constraints like M1 $\rightarrow$ A "Pronounce M1 as A."

Contrast tableau

		M1 → A	M1 → B	M2 → A	M2 → B
M1	A		-1		
	B	-1			
M2	A				-1
	B			-1	

Meanings are not apparent from surface forms, they must be inferred.

Meanings are not apparent from surface forms, they must be inferred.

The agents use Robust Interpretive Parsing (RIP; Tesar and Smolensky 2000, Boersma 2003, Jarosz 2013, Boersma and Pater 2014):

Meanings are not apparent from surface forms, they must be inferred.

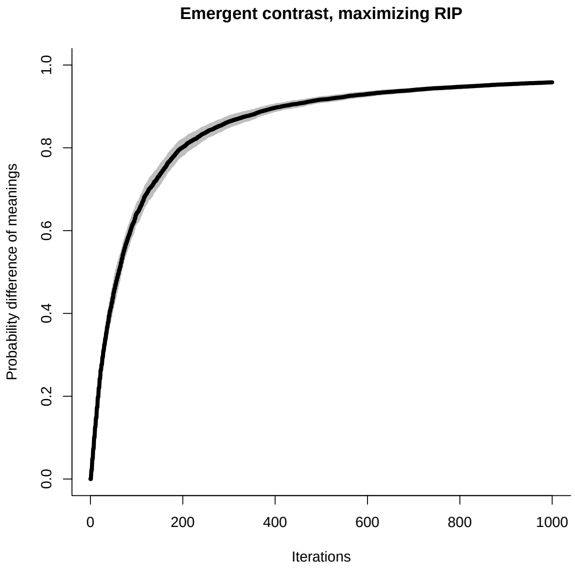
The agents use Robust Interpretive Parsing (RIP; Tesar and Smolensky 2000, Boersma 2003, Jarosz 2013, Boersma and Pater 2014):

Agents choose the meaning that they would most likely pronounce with the observed surface form.

Interpretation

Teacher Production:	M1	$\rightarrow$	$\boxed{a}$ / b
Interpretation:	a	$\rightarrow$	M1 / $\boxed{\text{M2}}$
Learner Production:	M2	$\rightarrow$	a / $\boxed{b}$

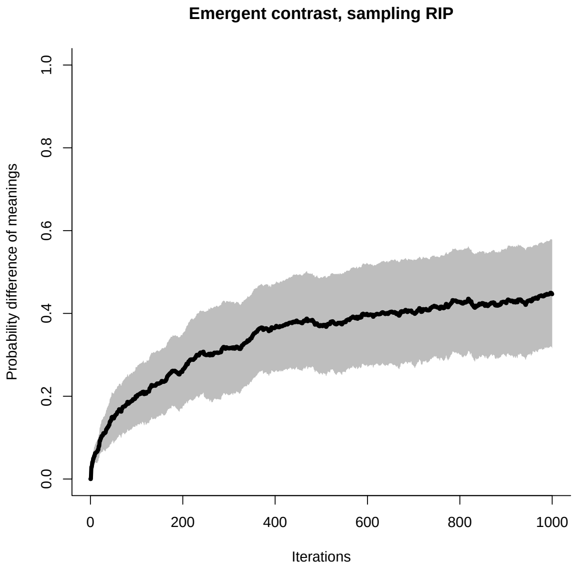
Update: Output is not a $\Rightarrow$ $\text{M2} \rightarrow a \uparrow$, $\text{M2} \rightarrow b \downarrow$

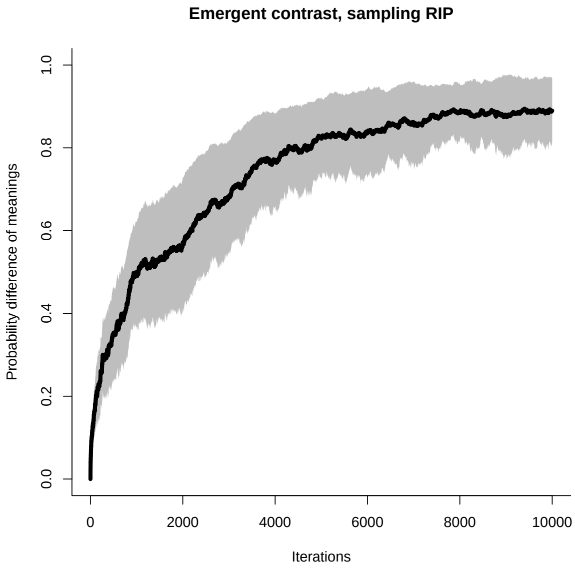

Emergent contrast, maximizing RIP
Probability difference of meanings
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
Iterations

Our assumption that RIP finds the *most* likely word accelerates contrast.

Errors in interpretation point to non-categorical probabilities—maximizing helps find these.

If we sample instead of maximizing, however, we still get this kind of trend.


Emergent contrast, sampling RIP
Probability difference of meanings
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
Iterations


Emergent contrast, sampling RIP
Probability difference of meanings
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000
Iterations

Similar patterns are found with iterative learning.

Similar patterns are found with iterative learning.

Similar pressures for categoricity $\rightarrow$ similar contrast effects.

Conclusions

We have shown language learners in a network tend towards stability with categorical grammars.

Conclusions

We have shown language learners in a network tend towards stability with categorical grammars.

This tendency emerges from interaction and transmission: Categorical patterns are those with the most reliability across generations and interactions.

This tendency addresses several possible issues with probabilistic models:

This tendency addresses several possible issues with probabilistic models:

- Why are languages more categorical than they could be?

This tendency addresses several possible issues with probabilistic models:

- Why are languages more categorical than they could be?
- How can categorical contrast emerge?

This tendency addresses several possible issues with probabilistic models:

- Why are languages more categorical than they could be?
- How can categorical contrast emerge?
- Why are gang effects not (seemingly) ubiquitous?

More broadly, this work reinforces the importance of viewing grammatical models in context:

1. We must consider learning models and their concomitant biases.
2. We must consider how these learning models interact to form typological patterns.

Thank you!

This material is based upon work supported by the National Science Foundation under Grant No. S121000000211 to the second author, Grants BCS-0813829 and BCS-424077 to the University of Massachusetts, and by the city of Paris under a Research in Paris fellowship to the third author.

We would also like to thank Lucien Carroll.