Syntactic universals in the lab: New methods and approaches

Jennifer Culbertson, University of Edinburgh

Advanced Core Training in Linguistics (ACTL), Summer 2015.
University College London.

Lecture 5: Models, theory and explanation

Overview

Formalizations of Universal 18, 20. Two cases of word order universals in which we have both typological and experimental data as well as well-specified theoretical models of the bias.

Universal 18: PCFG model[1]

[1] J. Culbertson, P. Smolensky, *Cognitive Science* **36**, 1468 (2012)

Inferring a PCFG given some data.

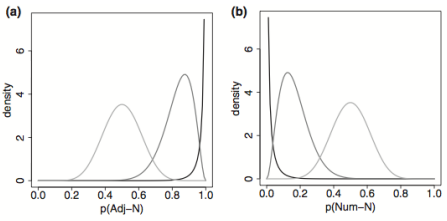
- What is a PCFG? It's a CFG with probabilities attached to productions
- Given a set of data, what probability should you infer?
- Binomial: (c=counts, t=total trials, p=probability)

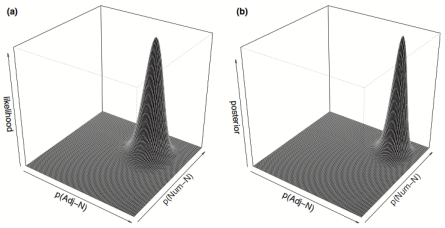
$$\text{binomial}(\text{c}|p) = \binom{t}{c} p^c (1-p)^{t-c}$$

- From the experiment: given 12 instances of Adj-N and 28 instances of N-Num...?

$$\text{binomial}(12|0.3) = 0.14 \qquad \text{binomial}(28|0.7) = 0.14 \qquad \text{binomial}(28|0.9) = 0.0003$$

- What if you have a regularization bias...?
- Combining (by multiplication) a beta distribution with a binomial...





Bayesian inference with PCFGs.

- We just did Bayesian inference!
- *Bayes rule:* contribution of input data and prior knowledge to learning

$$P(\text{Grammar}|\text{Data}) \propto P(\text{Data}|\text{Grammar})P(\text{Grammar})$$

- *Grammar:* probabilistic re-write rules p(Adj-N), p(Num-N)
- *Likelihood:* binomial probability of training counts given grammar
- *Prior:*
  - Beta distribution as formalization of regularization bias (add counts)
  - Multinomial weights on grammar types as pattern bias
- *What bias do learners have?* Start with flat prior (no bias), fit to behavioral data
- *Result:*
  - Strong regularization bias; $\alpha, \beta = (16.5, 0.001)$
  - Asymmetry among patterns; $\gamma = (0.62, 0.27, 0.0001, 0)$

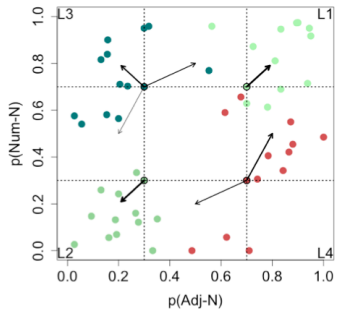


Figure 1: Basic result.

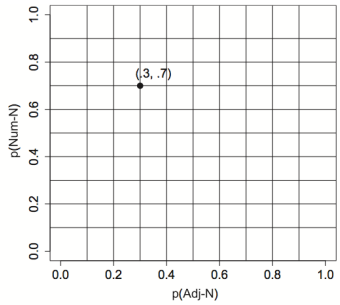


Figure 2: Grid approximation of space of PCFG grammars.

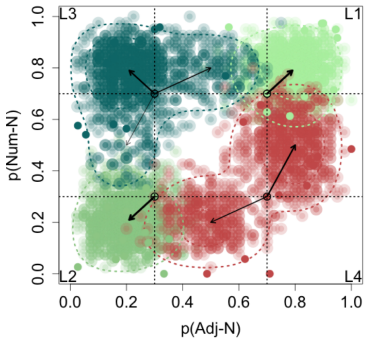
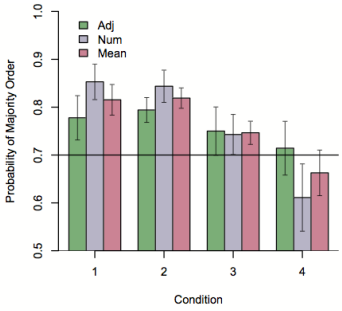


Figure 3: Predictive distribution of production grammars.

Universal 18: OT/PHG model[2]

[2] J. Culbertson, P. Smolensky, C. Wilson, *Topics in Cognitive Science* **5**, 392 (2013)

BASIC RESULTS. Regularization and quantitative difference in use of majority order for numeral phrases across condition.



a. ***Num position in L3***

{Num, N}		NUM-L	HEAD-R	HEAD-L
a. ☞	Num-N		*	
b.	N-Num	*!		*

b. ***Adj position in L3***

{Adj, N}		NUM-L	HEAD-R	HEAD-L
a.	Adj-N		*!	
b. ☞	N-Adj			*

FROM OT TO PROBABILISTIC HARMONIC GRAMMAR (AKA MAXENT).

- *OT constraints for Universal 18*
  - HEAD-L, HEAD-R, NUM-L
  - Allows L1, L2, L3, prohibits L4 (no way to derive it without NUM-R)
  - No way to predict harmonic > L3
- *Moving to PHG with bias*
  - HEAD-L, HEAD-R, NUM-L, NUM-R
  - Allow all possible patterns
  - Prior penalty for use of specific → favor harmonic
  - Combine with prior penalty for use of NUM-R → disfavor (Adj-N, N-Num)

Num position in L3 ($Z \equiv e^{-0.85} + e^{-1.7}$)

{Num, N}		1.7 NUM-L	.85 HEAD-R	0 HEAD-L	$H_G(x) = -\sum_k w_k C_k(x)$	$P(x\|G) = e^{H_G(x)}/Z$
a. ☞	Num-N		*		$-[1.7\cdot 0 + 0.85\cdot 1 + 0\cdot 0]$	$e^{-0.85}/Z = 70\%$
b.	N-Num	*!		*	$-[1.7\cdot 1 + 0.85\cdot 0 + 0\cdot 1]$	$e^{-1.7}/Z = 30\%$

- *How to formalize regularization?*
  - Bias over constraint weights (gaussian prior penalizing weights near zero)
  - Bias over resulting probabilities (beta distribution with asymmetric shape parameters, identical to PCFG model)
- *Results:*
  - Both models can capture average differences among conditions
  - Weight-based model captures asymmetry in numeral ordering better
  - Probability-based model is preferable on complexity grounds

OVERALL TAKE HOME MESSAGE

- The Bayesian view presented in both these treats underlying biases as soft
- All grammars (defined by the particular framework) are in principle part of the hypothesis space
- A priori biases make some less likely to be inferred
- Strong enough prior biases can be strong enough to prevent a particular grammar from being inferred (approximates hard/absolute constraints)

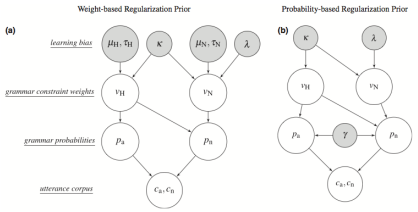


Figure 4: Alternative PHG models of regularization.

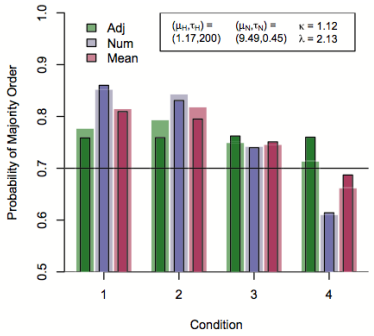


Figure 5: Model fit with weight-based prior.

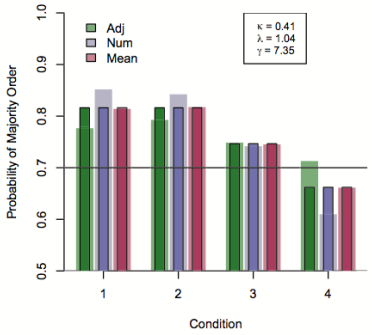


Figure 6: Model fit with probability-based prior.

Explanations and Universal 20

WHY? If U18 and U20 are the result of cognitive biases, what is the content of the biases? This brings us back to several of the larger issues we've been discussing:

- How to reconcile statistical generalizations with non-statistical theories?
- Are the biases part of the grammatical system or external to it?

THEORETICAL ACCOUNTS OF UNIVERSAL 20.

- Original and Hawkins' reformulation[3]

> *Universal 20.* When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite.

> …For those that follow, no predictions are made, though the most frequent order is the mirror image of the order for preceding modifiers. In no case does the adjective precede the head when the demonstrative or numeral follow.

- Cinque[4]

> Of the 24 mathematically possible orders of the four elements demonstrative, numeral, adjective, and noun, only 14 appear to be attested in the languages of the world. Some of these are unexpected under Greenberg's Universal 20. Here it is proposed that the actually attested orders, and none of the unattested ones, are derivable from a single, universal, order of Merge (Dem > Num > Adj > N) and from independent conditions on phrasal movement.

✓	a	Dem	Num	A	N	MANY	⦸	m	Dem	A	Num	N	zero
✓	b	Dem	Num	N	A	many	✓	n	Dem	A	N	Num	FEW
✓	c	Dem	N	Num	A	FEW	✓	o	Dem	N	A	Num	many
✓	d	N	Dem	Num	A	few	✓	p	N	Dem	A	Num	FEW
⦸	e	Num	Dem	A	N	zero	⦸	q	Num	A	Dem	N	zero
⦸	f	Num	Dem	N	A	zero	✓	r	Num	A	N	Dem	FEW
⦸	g	Num	N	Dem	A	zero	✓	s	Num	N	A	Dem	few
⦸	h	N	Num	Dem	A	zero	✓	t	N	Num	A	Dem	few
⦸	i	A	Dem	Num	N	zero	⦸	u	A	Num	Dem	N	zero
⦸	j	A	Dem	N	Num	zero	⦸	v	A	Num	N	Dem	zero
✓	k	A	N	Dem	Num	FEW	✓	w	A	N	Num	Dem	FEW
✓	l	N	A	Dem	Num	few	✓	x	N	A	Num	Dem	MANY

  - Linear Correspondence Axiom plus universal merge order
  - Movement of NP or XP containing it to c-commanding position

- Abels & Neeleman[5]
  - Any linearization of universal merge order plus leftward movement
  - Movement of sub-tree containing N to c-commanding position
- Steddy & Samek-Lodovici[6]: OT alignment constraints Dem-L, Num-L, Adj-L, N-L
- Issues[7]
  - Where to draw the line?
  - Do we actually want to predict that a type with frequency 1 is possible while an unattested type is impossible? Are we justified statistically?
  - Even worse, should we make this assumption at the cost of not explaining the huge frequency differences?
- Alternative explanation (with experimental support)
  - Universal semantic scope relations plus bias for isomorphic linearization
  - Bias against non-harmonic patterns
  - Bias against pre-nominal Adj (see also U18, where Adj-N, N-Num is WOW)

"It is a fact of logic that one cannot derive statistical predictions from a non-statistical theory. Therefore, providing an argument for any particular grammatical analysis on the basis of frequency information is an arduous task. The first step in constructing such an argument must consist in pairing the proposed analysis with a theory of markedness. The latter identifies grammatical structures as favored or disfavored, and thus generates statistical predictions. Unfortunately, little is known about this aspect of the language faculty. An independently motivated theory of markedness, which would allow us to test hypotheses about the grammar, is no more than a distant hope."

K. Abels, A. Neeleman, *Ms., University of Tromsø and University College London* (2006)

[3] J. A. Hawkins, *Word order universals* (Academic Press, New York, 1983)

[4] G. Cinque, *Linguistic Inquiry* **36**, 315 (2005)

[5] K. Abels, A. Neeleman, *Syntax* **15**, 25 (2012)

[6] S. Steddy, V. Samek-Lodovici, *Linguistic Inquiry* **42**, 445 (2011)

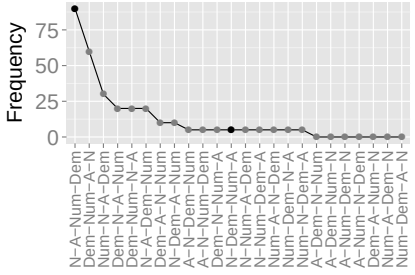


Figure 7: Distribution of 24 orders.

[7] M. Cysouw, *Linguistic Typology* **14**, 253 (2010)

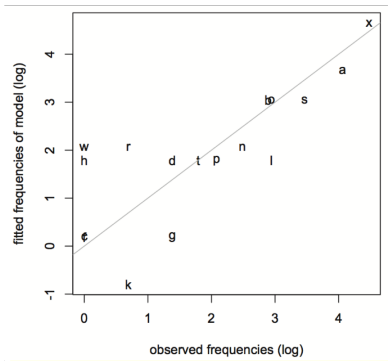


Figure 8: Fit of regression model with scope, harmony, N-Adj.