Computational Answers to Human Language Learning Questions

Lisa Pearl
University of Maryland
June 21, 2006

Road Map

Road Map

Introduction: Human Language Learning Questions

- Problem for human language learning research: what data do young learners learn *from*?
- Data learned from = **intake**
- Options:
  - Use all available data (which is noisy)
  - Use some subset of available data (which might be cleaner)

Introduction: Computational Answers

- *Can't* use traditional experimental methods since unnatural restriction of input to human learners for years has both logistical & ethical problems
- *Can* use computational simulation since we can easily restrict the input to virtual learners in any way we like and then see what the result is

Introduction: Virtual Learners

- Virtual learners instantiated with language learning model that allows probabilistic access of multiple structural options (Yang 2003, Bock & Kroch 1989)
- What virtual learners are learning: the probabilities used by mature speakers in the population for accessing the available structural options

Introduction: Proposals for Input Restriction

- Two proposals for restricting the intake of human learners to a subset of the available data
  - intake data is unambiguous
  - intake data is in main clauses

Introduction: Metric for Successful Language Learning

- How do we measure the effect of input restriction on human language learning?
- Use language change as a metric!

Introduction: Language Change As Metric

- Assume certain language changes occur because individual language learning is *imperfect* (Lightfoot, 1991) - population-level result is language change
- If simulated population with individuals using input restriction during learning can match the historically attested rate of language change, then this demonstrates successful language learning at the individual level

Road Map

I. Introduction

II. Human Language Learning Question: Restrictions on Input

III. How To Answer: Language Change Modeling

IV. The Model: Individual & Population-Level

V. Results and Conclusion

Restrictions on Input: Unambiguous Data

- Language has multiple options available for analyzing sentence structure - parameters (Chomsky, 1981)
- Each parameter can have several values that may be used cross-linguistically
- Proposal: learners use only unambiguous data, which can only be analyzed with one parameter value (Dresher 1999, Lightfoot 1999, Fodor, 1998)

Restrictions on Input: Unambiguous Data

- Advantage: Makes learning easier (no guesswork required for what parameter value should be chosen)
- Disadvantage: May be difficult to find (potential data sparseness problem)

Restrictions on Input: Main Clause Data

- Proposal: Human learners use only data in "simple" clauses, such as main clauses (also called degree-0 clauses) (Lightfoot 1991)

The clever boy thought that the giant was easy to fool.
[---------Degree-0-------]
[-------------Degree-1----------]

Restrictions on Input: Main Clause Data

- Advantage: may allow for the necessary *imperfect* learning that language change requires
- Disadvantage: when combined with unambiguous data proposal, compounds data sparseness problem

Restrictions on Input: Questions

Are these proposals (learning only from degree-0 unambiguous data) *viable* for accurately modeling human language learning?

If so, are they *necessary* to accurately model human language learning?

Road Map

I. Introduction

II. Human Language Learning Question: Restrictions on Input

III. How To Answer: Language Change Modeling

IV. The Model: Individual & Population-Level

V. Results and Conclusion

Language Change Modeling: Logic Recap

- Population-level result of language change comes from individual-level *imperfect* learning over time (Lightfoot 1991)
- If a simulated population with individuals using input restriction during learning can match the historically attested rate of language change, then this demonstrates successful language learning at the individual level

Language Change Modeling: Old English Language Change

- Shift in Old English between 1000 A.D. and 1200 A.D. from a strongly OV distribution to a strongly VO distribution (YCOE, PPCME2 historical corpora)

OV

he$_{\text{Subj}}$	hyne$_{\text{Obj}}$	gebidde$_{\text{TensedVerb}}$
He	*him*	*may-pray*

'He may pray (to) him'
(*Ælfric's Letter to Wulfsige,* 87.107, ~1075 A.D.)

Language Change Modeling: Old English Language Change

- Shift in Old English between 1000 A.D. and 1200 A.D. from a strongly OV distribution to a strongly VO distribution (YCOE, PPCME2 historical corpora)

VO

& [mid his stefne]$_{PP}$ he$_{Subj}$ awecð$_{Tensed\ Verb}$ deade$_{Obj}$ …
& *with his stem* *he* *awakened* *the-dead*
'And with his stem, he awakened the dead . . . '
(*James the Greater,* 30.31, ~1150 A.D.)

Language Change Modeling: Unambiguous OV/VO data

- Reasonable idea:
  - Unambiguous OV: …Object Verb…
  - Unambiguous VO: …Verb Object…

But other available structural options can interfere!

Language Change Modeling: Interfering Structural Options

- Verb-Second (V2) movement: the tensed Verb is moved to the second phrasal position and some other phrase moves to the first phrasal position (like modern German)

Example:
Subject TensedVerb t_{Subj} Object $t_{TensedVerb}$

This can produce "…Verb Object…" order, even if the underlying order is OV!

Language Change Modeling: Unambiguous OV/VO Data

Unambiguous OV data has the form (Lightfoot 1991)

- XP … Object Verb …

Ex: Subject Object Verb

- XP TensedVerb …Object Verb-Marker …

Ex: Subject TensedVerb Object Verb-Particle

Language Change Modeling: Unambiguous OV/VO Data

Unambiguous VO data has the form (Lightfoot 1991)

- XP1 XP2 …Verb Object …

Ex: Adverb Subject TensedVerb Object

- XP1 TensedVerb … Verb-Marker Object …

Ex: Subject TensedVerb NonTensedVerb Object

Language Change Modeling: Verb-Markers

- Verb-Markers are semantically associated with the Verb (such as verb-particles ('up'), nontensed verbs that are complements to the tensed verb ('shall perform'), negatives ('not'), and some closed-class adverbials ('never') (Lightfoot 1991)
- Verb-Markers are not usually subject to V2 movement - they mark the tensed verb's position before movement and allow more data to be considered unambiguous

Language Change Modeling: Ambiguous Data

- Nonetheless, Old English still has a large quantity of ambiguous data: 71-80% of degree-0 data is ambiguous, depending on the time period
- Could make data sparseness a problem for a learner that learns only from what is perceived as unambiguous data (question of *viability* for proposals)

Language Change Modeling: Potential For Success

- However, the very sparseness of the learner's intake could be an advantage: it allows the distribution of OV and VO utterances that the learner learns from to be different from the distribution that speakers use to generate those same utterances
- This allows *imperfect* learning in individuals, that will eventually leave to a population-level result: language change

Road Map

I. Introduction

II. Human Language Learning Question: Restrictions on Input

III. How To Answer: Language Change Modeling

IV. The Model: Individual & Population-Level

V. Results and Conclusion

Individual-Level: Probabilistic Access

- Individuals can access different structural options (OV vs. VO) probabilistically when producing utterances (Yang 2003, Bock & Kroch 1989)
- Languages like modern German and modern English access one option 100% of the time (OV for German, VO for English)
- Languages like Old English access both options

Individual-Level: Only One Option Accessed

- Probability of accessing VO option: p_{VO} (Probability of accessing OV option: $1 - p_{VO}$)
  - p_{VO} for modern German = 0.0
  - p_{VO} for modern English = 1.0

All unambiguous data will be unambiguous for only one option since speakers only ever use one option to generate their utterances

Individual-Level: Both Options Accessed

- $0.0 < p_{VO}$ for Old English < 1.0
- Learner is trying to determine the correct p_{VO}
- Some unambiguous data will be generated with the OV option and some with the VO option = conflicting unambiguous data

Individual-Level: Advantage

- Learner's initial p_{VO} = 0.5 (no bias for either option)
- Potential data sparseness problem: equal amounts of conflicting unambiguous data will cause learner to remain at 0.5. Only way to move away is to observe more unambiguous data for one option.
- How much more unambiguous data = option's advantage in the intake

Individual-Level: Data Sparseness

- Population checkpoints:
  - 1000 - 1150 A.D. = strongly OV (p_{VO} << 0.5)
  - 1200 A.D. = strongly VO (0.5 << p_{VO})

Must be sufficient advantage in the learner's intake for OV before 1150 A.D. and for VO after 1150 A.D. for the learner to converge on the appropriate p_{VO}.

Individual-Level: Advantage

- Old English OV advantage in degree-0 clauses (YCOE, PPCME2)

Time Period	D0 OV Advantage
1000 A.D.	**4.6%**
1000-1150 A.D.	**0.5%**
1200 A.D.	**-0.8%**

Individual-Level: Bayesian Learner

- Initial p_{VO} of 0.5 = learner expects the distribution of OV and VO utterances in the intake to be equally split
- Learner's expectation of utterances in the intake = binomial distribution centered around p_{VO}
- After each datum in the intake, learner updates p_{VO} by taking the MAP probability (sequence length = 1)

Individual-Level: Bayesian Learner

- If VO datum seen: $$p_{VO} = \frac{(p_{VOprev} * n + c)}{n + c}$$
- If OV datum seen: $$p_{VO} = \frac{(p_{VOprev} * n)}{n + c}$$

where *n* = number of utterances in intake (2000) and *c* = learner's confidence in input, scaled to make 0.0 <= p_{VO} <= 1.0

Individual-Level: Learning Algorithm

```
pVO = 0.5
IntakeCount = 0
while IntakeCount <= 2000
   get datum from input
   if datum = degree-0 unambiguous then
         update pVO using Bayesian updating
         IntakeCount = IntakeCount + 1
```

Population-Level: 1000 A.D. to 1200 A.D. Simulation

PopulationAgeRange = 0 to 60
PopulationSize = 18000
Time = 1000 A.D.
while Time <= 1200 A.D.
  Population members age 59-60 die off
  Remaining population members age 2 years
  New members are born
  New members use individual learning algorithm to
    set individual p_{VO}, input from rest of population
  Time = Time + 2

Model: Matching Historical Rate of Change

- To see if the simulated population is changing at the correct rate, we must derive the historically attested rate of change
- We do this by calculating the distribution of OV and VO access by speakers of the Old English population at various points in time

Model: Matching Historical Rate of Change

- To match the historically attested rate of change, the simulated population must have an average p_{VO} that matches the historically attested p_{VO} at various points in time

Time Period	(Initialization) 1000 A.D.	(Calibration) 1000-1150 A.D.	(Termination) 1200 A.D.
Average VO Access Value	0.23	0.31	0.75

Road Map

I. Introduction

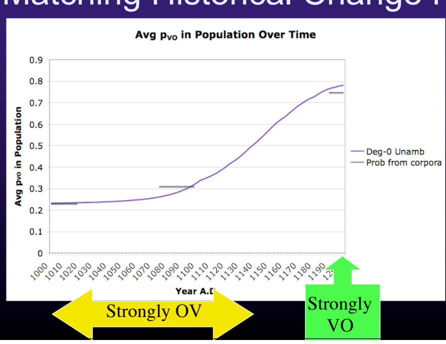
II. Human Language Learning Question: Restrictions on Input

III. How To Answer: Language Change Modeling

IV. The Model: Individual & Population-Level

V. Results and Conclusion

Learning Proposals are Viable: Matching Historical Change Rate



Learning Proposals are Necessary: Testing

- To see if the learning proposals are necessary, we can drop one or both of the restrictions on the individual learner's intake and see how a simulated population made up of such individuals would fare

Learning Proposals are Necessary: Drop Unambiguous Restriction

- Suppose we allow the learner to use ambiguous data, such as the "…Verb Object…" utterances for VO

VO Advantage in the learner's intake:

1000 A.D. : 13.8%
1000-1150 A.D.: 14.8%

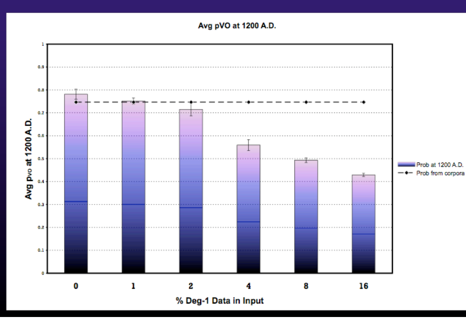
Impossible for population to remain strongly OV before 1150 A.D.

Learning Proposals are Necessary: Drop Degree-0 Restriction

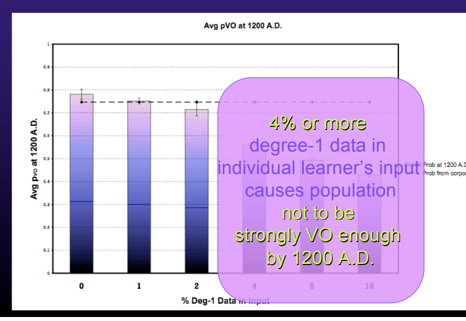
- Suppose we allow the learner to use degree-1 (embedded clause) data as well.
- The OV advantage for degree-1 data is much higher before 1150 A.D. than the degree-0 data OV advantage.

Time Period	D0 OV Advantage	D1 OV Advantage
1000 A.D.	**4.6%**	**29.9%**
1000-1150 A.D.	**0.5%**	**21.6%**

Learning Proposals are Necessary: Drop Degree-0 Restriction



Learning Proposals are Necessary: Drop Degree-0 Restriction



Learning Proposals are Necessary: Drop Degree-0 Restriction

- Estimates from modern English input to children suggest that 15-16% of it is degree-1 (CHILDES database, Sakas 2003)
- 4% or more degree-1 data causes population's rate of change to be too slow

Impossible for population without degree-0 restriction to match historically attested rate of change.

Learning Proposals are Necessary: Drop Both Restrictions

- Dropping unambiguous restriction causes population to change too quickly
- Dropping degree-0 restriction causes population to change too slowly

What if we drop both restrictions?

Learning Proposals are Necessary: Drop Both Restrictions

- VO advantage in learner's intake still makes change happen too quickly

  1000 A.D. degree-0: 13.8%
  1000 A.D. degree-1: -10.1%

- *Would need 56% degree-1 data in the input just to neutralize the VO advantage (over 3 times the amount estimated in modern English input to children)*

Conclusions

- Learning from a subset of the available data is both a viable and necessary method for human language learning
- Mathematical models and computational simulation can inform human language learning theory when traditional experimental methodology cannot

Thank you!

Thanks to…

Amy Weinberg	Charles Yang	David Lightfoot
Garrett Mitchener	LouAnn Gerken	Susan Goldin-Meadow
Susan Stevenson	William Sakas	Ted Briscoe
Norbert Hornstein	Stephen Crain	Rosalind Thornton
Tony Kroch	Beatrice Santorini	Ann Taylor
Susan Pintzuk	Philip Resnik	Cedric Boeckx
Michelle Hugue		

Audiences at…

ACL 2006	ICEHL 13
DIGS VIII	28th PLC
UMaryland CNL Lunch Talks	2003 UMaryland student conference.