The Input for Syntactic Acquisition: Solutions from Language Change Modeling

Lisa Pearl
University of Maryland
Psychocomp Workshop at ACL 2005
June 29th, 2005

Road Map

Road Map

I. Introduction

II. The Acquisition Proposals: Restrictions on the Learner

III. Old English Change

IV. The Population Model

V. Results and Conclusion

Introduction: Big Picture

- What we want to explore: the relation between human language acquisition and machine learning
  - similarities: probabilistic/statistical methods (psychologically plausible)
  - differences: acquisition is more finely tuned (small changes can have large effects over time)

- Data Sparseness: How much data does it take to get the job done? Especially if it's a learning system with very particular constraints on how learning takes place.

Introduction: Investigating the Input to Syntactic Acquisition

- traditional methods of investigating the input to syntactic acquisition won't work (ethical & logistical issues)

- having a population of simulated learners allows us to restrict the input any way we like and see the effects on the learners

Introduction: This Work

- simulated learners follow an acquisition model inspired by Yang (2003, 2000) - probabilistic access of multiple grammars

- use a population of these simulated learners to provide empirical support for two proposals from acquisition literature that are resource-sparing:
  - data must be unambiguous
  - data appears in simple clauses

Introduction: Using Language Change

• **language change** as a metric of successful **population-level acquisition**

• Logic: if simulated population with these input restrictions behaves as real population historically did, then simulated acquisition process is similar to real acquisition process

• Using data from Old English shift from a strongly Object-Verb order distribution to a strongly Verb-Object order distribution between 1000 and 1200 A.D.

Road Map

I. Introduction

II. The Acquisition Proposals: Restrictions on the Learner

III. Old English Change

IV. The Population Model

V. Results and Conclusion

Acquisition: Unambiguous Triggers

- adult grammar = a specific set of values for universal linguistic parameters (Chomsky 1981)
- **acquisition** = the process of determining those values
- **unambiguous trigger** (Dresher 1999, Lightfoot 1999, Fodor 1998): data that can *only* be parsed with one value

Acquisition: Unambiguous Triggers

- An unambiguous trigger for the value p1 of parameter *P* can be parsed only with value *p1* and not value *p2*, no matter what other parameter values (*a1 or a2, b1 or b2, c1 or c2,…*) are used
- Corresponds to exactly one parameter *P* and can only alter the value of *P (bypasses Credit Problem noted in Dresher (1999)*)

Unambiguous Triggers: The Good and the Bad

- *Advantage*: Resource sparing. No need to test 2^n grammars on every piece of data . Restriction to specific cues in the data that correspond to only **n** different parameters
- *Danger* : Data sparseness. May not be much data that is unambiguous.

Acquisition: Degree-0 Clauses

- *Proposal:* Children heed the data in **simple** clauses only (Lightfoot 1991, simple = **degree-0**)

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

Degree-0: The Good and the Bad

- *Advantage :* degree-0 data is messier = ***language change***
- *Danger :* compounding the potential data sparseness problem once combined with **unambiguous triggers**

Road Map

I. Introduction

II. The Acquisition Proposals: Restrictions on the Learner

III. Old English Change

IV. The Population Model

V. Results and Conclusion

Old English Change

• shift in Old English word order between 1000 and 1200 A.D. from a strongly **OV** distribution to a strongly **VO** distribution (YCOE, PPCME2)

• **OV** order (*Beowulf* 625)
he Gode þancode
he God thanked
"He thanked God."

• **VO** order (*Alcuin De virtutibus et vitiis*, 83.59)
heo clænsað þa sawle þæs rædendan
he purified the souls [the advising]-Gen
"He purified the souls of the advising ones."

Old English Change: Degree-0 Unambiguous Triggers

• **OV** order unambiguous trigger
$_{VP}$[Object Verb] or $_{VP}$[Object Verb-Marker]

• **VO** order unambiguous trigger
$_{VP}$[Verb Object] or $_{VP}$[Verb-Marker Object]

• **appropriate O-language order + unambiguous parse**

Old English Language Change: Ambiguous Triggers

• **ambiguous utterances**

Subject	Verb	Object
heo	clænsað	þa sawle þæs rædendan
he	purified	the souls [the advising]-Gen

"He purified the souls of the advising ones."

*Subject **Verb Object** t_{Verb}* (***OV*** *+ V2)*

OR

*Subject **Verb** t_{Verb} **Object*** (***VO*** *+ V2)*

Old English: Verb Markers

• Sometimes, a Verb-Marker (particles ('*up*'), negatives ('*not*'), some closed class adverbials ('*never*'), non-finite complements ('shall…*perform*') (Lightfoot 1991)) will be next to the Object so the utterance is unambiguous.

• underlying **OV** order
þa ahof Paulus his heafod up
then lifted Paul his head up

• underlying **VO** order
þa ahof Paulus up his heafod
then lifted Paul up his head

Old English: Ambiguous Input

(YCOE and PPCME2 Corpora)
% Ambiguous Utterances

Time Period	Degree-0 % Ambiguous	Degree-1 % Ambiguous
1000 A.D.	76%	28%
1000 - 1150 A.D.	80%	25%
1200 A.D.	71%	10%

Old English Change: Degree-0 Unambiguous Triggers

- Children learn from the degree-0 unambiguous trigger distribution, which is *different* from the distribution in the adult population.
- Children can **"misconverge".**
- Result: These small changes spread through an exponentially growing population until their effect manifests as a sharper population-level language change.

Road Map

I. Introduction

II. The Acquisition Proposals: Restrictions on the Learner

III. Old English Change

IV. The Population Model

V. Results and Conclusion

The Model: Foundations

- **Grammars can coexist**
  - during individual acquisition (Clark & Roberts 1993)
  - within a population over time (Pintzuk 2002, among others)
  - OV and VO grammar
- Population-level change is the result of **individual level "misconvergences"** (Niyogi & Berwick, 1997, 1996, 1995)
  - **individual linguistic behavior is a statistical distribution of multiple grammars** (Yang 2003, 2000)

The Model: Multiple Grammars

- individual can access *g* grammars with some probability p_g allotted to each (Yang 2002, 2000)

•system with single grammar:

$g = 1 \quad p_1 = 1$

- all unambiguous triggers are from this grammar

- system with multiple grammars:

$g = n \quad p_1 = P(g_1), p_2 = P(g_2), \ldots, p_n = P(g_n)$

$\Sigma_1^n p_n = 1$

- each grammar leaves some unambiguous triggers

The Model: Trigger Advantage

- **triggers for conflicting grammars:** what matters is ***how many <u>more</u>*** unambiguous triggers one grammar has than the other in the input
- **advantage:** how many more unambiguous triggers in input
- **Old English advantage in input. (YCOE, PPCME2)**

Time Period	D0 OV Advantage	*D1 OV Advantage*
1000 A.D.	**1.6%**	***11.3%***
1000-1150 A.D.	**0.2%**	***7.7%***
1200 A.D.	**-0.4%**	***-19.1%***

The Model: Being On Time

- Metric: Old English = strongly **OV** from **1000-1150 A.D.** and then more strongly **VO** by **1200 A.D.** (YCOE, PPCME2)
- Is the **restricted set** of degree-0 unambiguous triggers ***sufficient*** to get this change to emerge?
- Are the **restrictions *necessary*** to be on time?
  - Relax Unambiguous Triggers Restriction: allow ***ambiguous*** triggers in?
    - become strongly VO too soon?
  - Relax Degree-0 Restriction: allow **degree-1** data in?
    - become strongly VO too late?

The Model: Individual Acquisition Implementation

- **probabilistic access function** of binary parameter values (Bock & Kroch 1989)
  Example - OV/VO: 70% OV order and 30% VO order
- this distribution ≠ O-language unambiguous trigger distribution (Ambiguous Input in O-language)
- **acquisition effect**: child ***must*** acquire probabilistic access function to account for the conflicting triggers in the input

The Model: Individual Acquisition Implementation

- OV = 0.0 value and VO = 1.0 value
- **probabilistic access function value** = what percentage is VO
  Ex: 30% VO, 70% OV = .30
- no default preference = initial setting of 0.5
- two methods that make model more psychologically plausible by relativizing data's influence on learner, based on the prior confidence value for the appropriate grammar (the current VO access value)

The Model: Noise Filter

- **Noise Filter**
  - designed to separate "signal" from "noise"
  - "noise" depends on current VO access value
  - unambiguous triggers from minority grammar more likely to be construed as "noise"

Probabilistic value of VO access = 0.3
If next unambiguous trigger = VO
= **noise** with **70%** chance and ignored
= **signal** with **30%** chance and heeded
If next unambiguous trigger = OV
= **noise** with **30%** chance and ignored
= **signal** with **70%** chance and heeded

The Model: Batch Learner Method

- **Batch Learner Method**
  - how many triggers to alter current value
  - how many depends on current VO access value
  - the more the grammar is in the majority, the smaller its "batch" of triggers has to be

VO Value	OV Triggers Required	VO Triggers Required
0.0-0.2	1	5
0.2-0.4	2	4
0.4-0.6	3	3
0.6-0.8	4	2
0.8-1.0	5	1

The Model: Batch Learner Method

Probabilistic value of VO access = 0.3
If next unambiguous trigger = VO
if **4th VO** trigger seen,
alter value of VO access towards VO
If next unambiguous trigger = OV
if **2nd OV** trigger seen,
alter value of VO access towards OV

The Model: Individual Acquisition Process

Initial setting = 0.5;
While in Critical Period
Get one input utterance from linguistic environment created by rest of the population;
If utterance contains unambiguous trigger
If utterance passes through Noise Filter;
If enough triggers seen to make batch
Alter current value;

The Model: Population Process

Initialization:
Population Age Range = 0 to 60;
Initial Population Size = 18,000;
Initialize all members to initial VO access value;
At 1000 A.D., and every 2 years after until 1200 A.D.
Members age 59 to 60 die; rest age 2 years
Create new members age 0 to 1;
New individuals use individual acquisition process to set VO access value

The Model: Population VO Access Values

• VO access value = distribution of OV and VO utterances *before* trigger destruction causes some utterances to become ambiguous

• historical data = distribution *after* trigger destruction has caused some utterances to become ambiguous

The Model: Ambiguous Triggers

(YCOE and PPCME2 Corpora)
% Ambiguous Utterances

Time Period	Degree-0 % Ambiguous	Degree-1 % Ambiguous
1000 A.D.	76%	28%
1000 - 1150 A.D.	80%	25%
1200 A.D.	71%	10%

The Model: Determining Distributions

• **Degree-0** clauses have more trigger destruction and a more distorted OV/VO distribution than **degree-1 clauses**

• use the difference in distortion between **degree-0** and **degree-1** distribution to estimate the difference in distortion between **degree-1** and **underlying** distribution

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

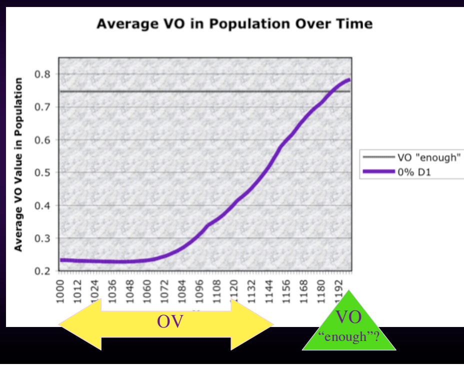
II. Language Change: Old English OV "Loss"

III. Acquisition Theory: Restrictions on the Learner

IV. The Population Model

V. Results and Conclusion

Results: Sufficiency



Results: Necessity of Unambiguous Triggers

- **allow ambiguous utterances with the right word order**
- surface order **VO** trigger (with ambiguous parse)

Subject Verb t_{Verb} **Object** t_{Verb} (**OV or VO + V2**)

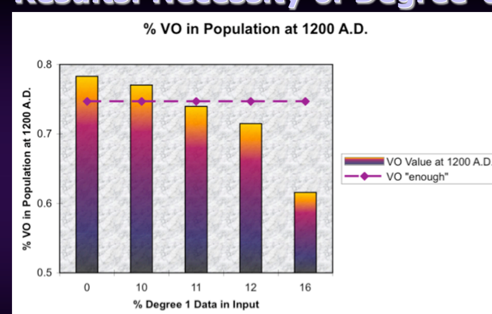
Time Period	Degree-0 VO Advantage
1000 A.D.	4.8%
1000-1150 A.D.	5.5%
1200 A.D.	8.5 %

- **VO too soon**

Results: Necessity of Degree-0

- Recall: Degree-1 data has a stronger OV advantage (stronger pull towards OV)
- If enough degree-1 data in input, population may not shift to a strong-enough VO distribution by 1200 A.D.
- estimates from modern children's corpora: input from adults consists of ~16% degree-1 data
  - *Is this too much?*

Results: Necessity of Degree-0



- **VO too late**

Conclusions: Acquisition and Change

- **language change + quantified model = novel testing ground for acquisition proposals**
- **future work: more sophisticated individual acquisition model involving Bayesian updating of a probabilistic distribution as well as a more realistic length of critical period**
- **future work: use the model to test out unambiguous triggers on other types of language change, such as Middle English Verb-Second Movement loss (Yang 2003, Lightfoot 1999, among others)**

Conclusions: Bigger Picture

- **probabilistic learning is psychologically realistic, and this is similar to the way machine learning is implemented now**
- **small errors during learning can add up over time (and cause language change), so human learning seems to be more sensitive and easily perturbed by "noise" during learning than machine learning**
- **data sparseness: human learning can take place even when the input is very restricted. Input serves to tune a system that already comes equipped with a lot of information about the way language works.**

Much thanks to Amy Weinberg, Charles Yang, Garrett Mitchener, David Lightfoot, Norbert Hornstein, Stephen Crain, Rosalind Thornton, Tony Kroch, Beatrice Santorini, Ann Taylor, Susan Pintzuk, Philip Resnik, Partha Niyogi, Cedric Boeckx, and Michelle Hugue.

Also, I am incredibly grateful for the parsed corpora of Old and Middle English made available through the folks at UPenn. Without such corpora, this work would have been nearly impossible.