Psych 215L: Language Acquisition

Lecture 17
Complex Systems

Complex Linguistic Systems

What is the generative system that creates the observed (structured) data of language (ex: metrical phonology)?

Observable data: stress contour EMphasis

Complex Linguistic Systems

What is the generative system that creates the observed (structured) data of language (ex: metrical phonology)?

Observable data: stress contour EMphasis

(S S) S
EM pha sis

(H L) H
EM pha sis

(H L L)
EM pha sis

(S S S)
EM pha sis

The Hypothesis Space

Which syllable of a larger unit is stressed?
{Leftmost, Rightmost}

Are all syllables included?
{Yes, No-not leftmost, No-not rightmost}

Are syllables differentiated?
{No, Yes-2 distinctions, Yes-3 distinctions}

Linguistic parameters = finite (if large) hypothesis space of possible grammars

Modeling learnability vs. modeling acquirability

- Modeling learnability
  - "Can it be learned at all by a simulated learner?"
  - "ideal", "rational", or "computational-level" learners
  - what is possible to learn

- Modeling acquirability (Johnson 2004)
  - "Can it be learned by a simulated learner that is constrained in the ways humans are constrained?"
  - more "realistic" or "cognitively inspired" learners
  - what is possible to learn if you're human

Knowledge Representation Motivations

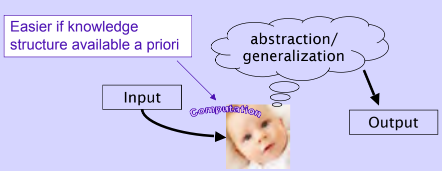
- One traditional motivation for proposals of knowledge representation (such as parameters): The knowledge representation helps explain the constrained variation observed in adult linguistic knowledge across the languages of the world

Argument from constrained cross-linguistic variation

Knowledge Representation Motivations

- Another (sometimes implicit) motivation for proposals of knowledge representation: Having this knowledge representation pre-specified allows children to acquire the right generalizations from the data as quickly as they seem to do

Argument from acquisition



Knowledge Representation Motivations

- Another (sometimes implicit) motivation for proposals of knowledge representation: Having this knowledge representation pre-specified allows children to acquire the right generalizations from the data quickly as they seem to do

Argument from acquisition

Pearl 2008, 2009, 2011

- Using computational methods and available empirical data, we can quantify this argument and explicitly test different proposals for knowledge representation
- At the same time, we can explore how acquisition could proceed if children were using these different knowledge representations

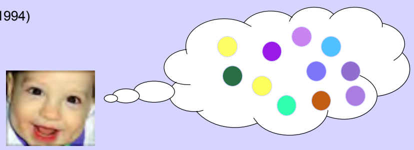
Learning Parametric Linguistic Systems

Linguistic parameters give the benefit of a finite hypothesis space. Still, the hypothesis space can be quite large.

For example, assuming there are *n* binary parameters, there are 2^n core grammars to choose from.

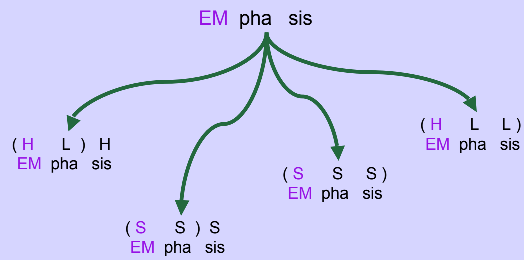
Exponentially growing hypothesis space

(Clark 1994)



Learning Parametric Linguistic Systems

Also, data are often *ambiguous* between competing hypotheses, since multiple grammars can account for the same data point.



Parametric Metrical Phonology

Metrical phonology:
What tells you to put the EMphasis on a particular SYLlable

Process speakers use:

Basic input unit: syllables — em pha sis

↓

Larger units formed: metrical feet — (em pha) sis

The way these are formed varies from language to language.

↓

Stress assigned within metrical feet — (EM pha) sis

The way this is done also varies from language to language.

↓

Observable Data: stress contour of word — EMphasis

Parametric Metrical Phonology

Metrical phonology:
What tells you to put the EMphasis on a particular SYLlable

Process speakers use:

Basic input unit: syllables — em pha sis

↓

Larger units formed: metrical feet — (em pha) sis

The way these are formed varies from language to language.

↓

Stress assigned within metrical feet — (EM pha) sis

The way this is done also varies from language to language.

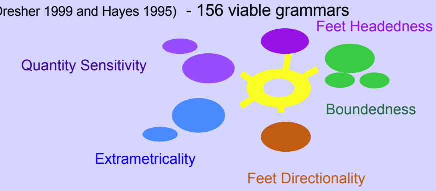
↓

Observable Data: stress contour of word — EMphasis

system parameters of variation - to be determined by learner from available data

Parametric Metrical Phonology

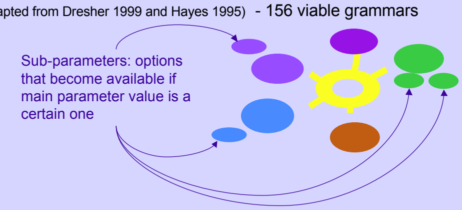
Metrical phonology system here: 5 main parameters, 4 sub-parameters
(adapted from Dresher 1999 and Hayes 1995) - 156 viable grammars



All combine to generate stress contour output

Parametric Metrical Phonology

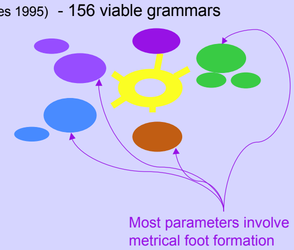
Metrical phonology system here: 5 main parameters, 4 sub-parameters
(adapted from Dresher 1999 and Hayes 1995) - 156 viable grammars



All combine to generate stress contour output

Parametric Metrical Phonology

Metrical phonology system here: 5 main parameters, 4 sub-parameters
(adapted from Dresher 1999 and Hayes 1995) - 156 viable grammars



All combine to generate stress contour output

A Brief Tour of Parametric Metrical Phonology

Are syllables differentiated?

No: system is quantity-insensitive (QI)

S	S	S
CVV	CV	CCVC
lu	di	crous

A Brief Tour of Parametric Metrical Phonology

Are syllables differentiated?

No: system is quantity-insensitive (QI)

S	S	S
CVV	CV	CCVC
lu	di	crous

Yes: system is quantity-sensitive (QS)

Only allowed method: differ by rime weight

krəs
crous
Syllable
onset — rime
kr
nucleus — coda
ə — s

CVV	CV	CCVC
lu	di	crous

A Brief Tour of Parametric Metrical Phonology

Are syllables differentiated?

No: system is quantity-insensitive (QI)

S	S	S
CVV	CV	CCVC
lu	di	crous

Yes: system is quantity-sensitive (QS)

Only allowed method: differ by rime weight ← narrowing of hypothesis space
Only allowed number of divisions: 2
Heavy vs. Light

VV always Heavy
V always Light

Option 1: VC Heavy (QS-VC-H)

H	L	H
CVV	CV	CCVC
lu	di	crous

Option 2: VC Light (QS-VC-L)

H	L	L
CVV	CV	CCVC
lu	di	crous

A Brief Tour of Parametric Metrical Phonology

Are all syllables included in metrical feet?

Yes: system has no extrametricality (Em-None)

(	...	)
L	L	H
VC	VC	VV
af	ter	**noon**

A Brief Tour of Parametric Metrical Phonology

Are all syllables included in metrical feet?

Yes: system has no extrametricality (Em-None)

(	...	)
L	L	H
VC	VC	VV
af	ter	**noon**

No: system has extrametricality (Em-Some)

Only allowed # of exclusions: 1
Only allowed exclusions: ← narrowing of hypothesis space
Leftmost or Rightmost syllable

A Brief Tour of Parametric Metrical Phonology

Are all syllables included in metrical feet?

(...)
L L H
VC VC VV
af ter **noon**

Yes: system has no extrametricality (Em-None)

No: system has extrametricality (Em-Some)

Only allowed # of exclusions: 1
Only allowed exclusions:
Leftmost or Rightmost syllable

narrowing of hypothesis space

Leftmost syllable excluded: Em-Left

(...)
L H L
V VC V
a gen da

Rightmost syllable excluded: Em-Right

(...)
H L H
VV V VC
lu di crous

A Brief Tour of Parametric Metrical Phonology

What direction are metrical feet constructed?

Two logical options

From the left:
Metrical feet are constructed from the left edge of the word (**Ft Dir Left**)

(→
H L H
VV V VC
lu di crous

From the right:
Metrical feet are constructed from the right edge of the word (Ft Dir Right)

←)
H L H
VV V VC
lu di crous

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

narrowing of hypothesis space

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

Ft Dir Left →

L L L H L
(L L L H L
(L L L)(H L
(L L L)(H L)

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

Ft Dir Left →

(L L L)(H L)

← Ft Dir Right

L L L H L

L L L H L)

L L L H) (L)

(L L L H) (L)

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

Ft Dir Left →

(L L L)(H L)

← Ft Dir Right

(L L L H) (L)

Ft Dir Left/Right

(L L L L L

(L L L L L)

S S S S S)

(S S S S S)

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space

---

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space

Ft Dir Left →

2 units per foot (Bounded-2)

x x x x

(x x) (x x

(x x) (x x)

3 units per foot (Bounded-3)

x x x x

(x x x) (x

(x x x) (x)

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space

(x x) (x x) B-2
(x x x) (x) B-3

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space
The counting units are restricted to 2 options: syllables or moras.

(x x) (x x) B-2
(x x x) (x) B-3

Ft Dir Left
Bounded-2
→
x x

(H L)(L H)

(L L) (L H) ← Count by syllables (Bounded-Syllabic)

(S S) (S S)

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space
The counting units are restricted to 2 options: syllables or moras.

(x x) (x x) B-2
(x x x) (x) B-3

Count by syllables (Bounded-Syllabic)

(H L)(L H)

Ft Dir Left
Bounded-2
→
x x

Count by moras (Bounded-Moraic)

xx x x xx
H L L H

(H) (L L) (H)

← Moras (unit of weight):
H = 2 moras **xx**
L = 1 mora **x**

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

(L L L H) (L)
(L L L)(H L)
(L L L L L)
(S S S S S)

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3. ← narrowing of hypothesis space
The counting units are restricted to 2 options: syllables or moras.

(x x) (x x) B-2
(x x x) (x) B-3

Count by syllables (Bounded-Syllabic)

Ft Dir Left
Bounded-2

Count by moras (Bounded-Moraic)

(H L)(L H) ←→ (H) (L L) (H)

compare

A Brief Tour of Parametric Metrical Phonology

Within a metrical foot, which syllable is stressed?

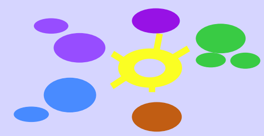
Two options, hypothesis space restriction

Leftmost:
Stress the leftmost syllable (Ft Hd Left) (H) (L L) (H)

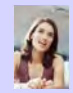
(H) (L L) (H)

Rightmost:
Stress the rightmost syllable (Ft Hd Right) (H) (L L) (H)

Generating a Stress Contour



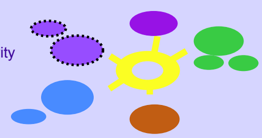
Process speaker uses to generate stress contour



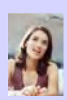
VC	CV	CVC
em	pha	sis

Generating a Stress Contour

Quantity Sensitivity



Process speaker uses to generate stress contour



Are syllables differentiated?

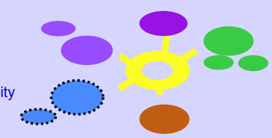
Yes - by rime.

VC & VV syllables are Heavy, V syllables are Light.

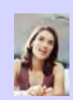
H	L	H
VC	CV	CVC
em	pha	sis

Generating a Stress Contour

Extrametricality



Process speaker uses to generate stress contour



Are any syllables extrametrical?

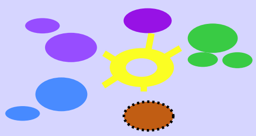
Yes.

Rightmost syllable is not included in metrical foot.

(…)

H	L	H
VC	CV	CVC
em	pha	sis

Generating a Stress Contour



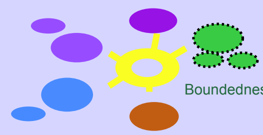
Process speaker uses to generate stress contour

Which direction are feet constructed from?

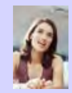
From the right.

H L) H
VC CV CVC
em pha sis

Generating a Stress Contour



Process speaker uses to generate stress contour



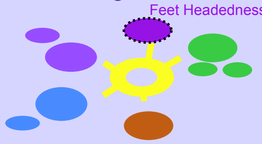
Are feet unrestricted in size?

No.

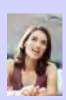
2 syllables per foot.

(H L) H
VC CV CVC
em pha sis

Generating a Stress Contour



Process speaker uses to generate stress contour

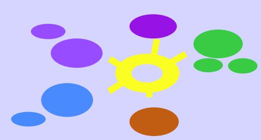


Which syllable of the foot is stressed?

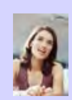
Leftmost.

(H L) H
VC CV CVC
em pha sis

Generating a Stress Contour

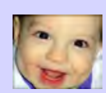


Process speaker uses to generate stress contour



Learner's task: Figure out which parameter values were used to generate this contour.

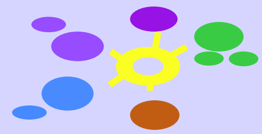
(H L) H
VC CV CVC
EM pha sis



Non-trivial case study: English

- Non-trivial because there are many data that are ambiguous for which parameter value or constraint ranking they implicate
- Non-trivial because there are many irregularities
  - Analysis of child-directed speech (8 -15 months) from Brent corpus (Brent & Siskind 2001) from CHILDES (MacWhinney 2000): 504084 tokens, 7390 types
  - For words with 2 or more syllables:
    - 174 unique syllable-rime type combinations (ex: closed-closed (VC VC))
    - 85 of these 174 have more than one stress contour associated with them (unresolvable): no one grammar can cover all the data
      - Ex for VC VC type: *her SELF*
        AN swer
        SOME WHERE

Cognitively inspired learners using parameters



- Target state = grammar for English (Halle & Vergnaud 1987, Dresher & Kaye 1990, Dresher 1999) derived from cross-linguistic variation and adult linguistic knowledge: quantity sensitive, VC syllables are heavy, rightmost syllable is extrametrical, feet are constructed from the right, feet are 2 syllables, feet are headed on the left

Premise: This is the grammar that best describes the systematic data of English, even if there are exceptions.

Biased learner, using only unambiguous data

Pearl (2008): Success is guaranteed as long as the parameters are learned in a particular order.

However…this requires the learner to identify unambiguous data and know/derive the appropriate parameter-setting order, which may not be trivial.

So…is this selective learning bias really necessary? How well do unbiased learners do?

Two psychologically plausible probabilistic update procedures

Yang (2002)

Naïve Parameter Learner (NParLearner)

Probabilistic generation & testing of grammars. (incremental)
Hypothesis update: Linear reward-penalty
(Bush & Mosteller 1951)

Two psychologically plausible probabilistic update procedures

Yang (2002)

Naïve Parameter Learner (NParLearner)

Probabilistic generation & testing of grammars. (incremental)
Hypothesis update: Linear reward-penalty
(Bush & Mosteller 1951)

MAP Bayesian Learner (BayesLearner)

Probabilistic generation & testing of grammars. (incremental)
Hypothesis update: Bayesian updating
(Chew 1971: binomial distribution)

Cognitively inspired learners using parameters

- Learner's algorithm:
  - Probabilistic generation and testing of parameter value combinations [grammars] (Yang 2002)
  - For each parameter, the learner associates a probability with each of the competing parameter values. Initially all values are equiprobable.
    - Ex: Quantity Sensitivity

      Value 1: Quantity Sensitive (0.5)   Value 2: Quantity Insensitive (0.5)
  - For each data point, a grammar is probabilistically generated, based on the probabilities associated with each parameter's values.

OCtopus

Cognitively inspired learners using parameters

- The selected grammar is then used to generate a stress contour, based on the syllable structure of the word.

VC	V	VC
oc	to	pus

- If the generated contour matches the observed contour, all participating parameter values are rewarded. If it mismatches, all values are punished.

OCtopus

OC to pus   oc TO pus

- Over time (as measured in data points encountered), the probability associated with a parameter value will approach either 1.0 or 0.0, based on rewards and/or punishments. Once the probability is close enough, the learner sets the appropriate parameter value.

Probabilistic learning for English

Probabilistic generation and testing of grammars (Yang 2002)

Update parameter value probabilities

NParLearner (Yang 2002): Linear Reward-Penalty

Learning rate γ:
small = small changes
large = large changes

Parameter values v1 vs. v2

reward v1	punish v1
$p_{v1} = p_{v1} + \gamma(1 - p_{v1})$	$p_{v1} = (1 - \gamma)p_{v1}$
$p_{v2} = 1 - p_{v1}$	$p_{v2} = 1 - p_{v1}$

Probabilistic learning for English

Probabilistic generation and testing of grammars (Yang 2002)

Update parameter value probabilities

NParLearner (Yang 2002): Linear Reward-Penalty

Learning rate γ:
small = small changes
large = large changes

Parameter values v1 vs. v2

$p_{v1} = p_{v1} + \gamma(1 - p_{v1})$ $\quad$ $p_{v1} = (1-\gamma)p_{v1}$

$p_{v2} = 1 - p_{v1}$ $\quad$ $p_{v2} = 1 - p_{v1}$

reward v1 $\quad$ punish v1

BayesLearner: Bayesian update of binomial distribution (Chew 1971)

Parameters α, β:

$\alpha = \beta$: initial bias at p = 0.5
α, β < 1: initial bias toward endpoints (p = 0.0, 1.0)

here: $\alpha = \beta = 0.5$

Parameter value v1

$$p_v = \frac{\alpha + 1 + \textit{successes}}{\alpha + \beta + 2 + \textit{total data seen}}$$

reward: success + 1 $\quad$ punish: success + 0

Probabilistic learning for English: Modifications

Probabilistic generation and testing of grammars (Yang 2002)

Update parameter value probabilities

Count-learning: smooth out some of the irregularities in the data, better deal with complex systems (Yang 2002)

Implementation (Yang 2002):
Matching contour = increase parameter value's batch counter by 1
Mismatching contour = decrease parameter value's batch counter by 1

Invoke update procedure (Linear Reward-Penalty or Bayesian Updating) when count limit *c* is reached.

Probabilistic learning for English: Modifications

Probabilistic generation and testing of grammars (Yang 2002)

Update parameter value probabilities + Count Learning

NParLearner (Yang 2002): Linear Reward-Penalty

Invoke when the batch counter for p_{v1} or p_{v2} equals *c*.

Parameter values v1 vs. v2

$p_{v1} = p_{v1} + \gamma(1 - p_{v1})$ $\quad$ $p_{v1} = (1-\gamma)p_{v1}$

$p_{v2} = 1 - p_{v1}$ $\quad$ $p_{v2} = 1 - p_{v1}$

reward v1 $\quad$ punish v1

BayesLearner: Bayesian update of binomial distribution (Chew 1971)

Invoke when the batch counter for p_{v1} or p_{v2} equals *c*.

Note: total data seen + 1

Parameter value v1

$$p_v = \frac{\alpha + 1 + \textit{successes}}{\alpha + \beta + 2 + \textit{total data seen}}$$

reward: success + 1 $\quad$ punish: success + 0

Cognitively inspired learners using parameters

Empirical grounding

- Learner's input based on the number of words likely to be heard on average in a 6 month period: 1,666,667. (Akhtar et al. (2004), citing Hart & Risley (1995)).

- Input distributions derived from child-directed speech distributions.
  - Brent corpus (Brent & Siskind 2001): 8 - 15 months
  - Child's syllabification of words: MRC Psycholinguistics Database (Wilson 1988)
  - Associated stress contour: CALLHOME American English Lexicon (Canavan et al. 1997)

Cognitively inspired learners using parameters

- Learner's algorithm:
  - Incremental update: words are processed one at a time, as they are encountered. (Assumes word segmentation is operational. Jusczyk, Houston, & Newsome (1999) suggests that 7-month-olds can segment some words successfully.)
  - Words are divided into syllables, with syllable rime identified as closed (VC), short (V), long (VV), or superlong (VVC). Jusczyk, Goodman, & Baumann (1999) and Turk, Jusczyk, & Gerken (1995) suggest young infants are sensitive to syllables and properties of syllable structure.
  - Sub-parameters are not set until the main parameter is set. This is based on the idea that children only consider information about a sub-parameter if they have to.

Probabilistic learning for English

Goal: Converge on English values after learning period is over

Learning Period Length: 1,666,667 words (based on estimates of words heard in a 6 month period, using Akhtar et al. (2004)).

QS, QSVCH, Em-Some, Em-Right, Ft Dir Right, Bounded, Bounded-2, Bounded-Syllabic, Ft Hd Left

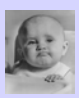
Probabilistic learning for English

Goal: Converge on English values after learning period is over

Learning Period Length: 1,666,667 words (based on estimates of words heard in a 6 month period, using Akhtar et al. (2004)).

QS, QSVCH, Em-Some, Em-Right, Ft Dir Right, Bounded, Bounded-2, Bounded-Syllabic, Ft Hd Left

Model	Success rate (1000 runs)
NParLearner, γ = .001, .0025, .01, .025	0.0%
BayesLearner	0.0%



Examples of incorrect target grammars

NParLearner:

Em-None, Ft Hd Left, Unb, Ft Dir Left, QI

QS, Em-None, QSVCH, Ft Dir Rt, Ft Hd Left, B-Mor, Bounded, Bounded-2

BayesLearner:

QS, Em-Some, Em-Right, QSVCH, Ft Hd Left, Ft Dir Rt, Unb

Bounded, B-Syl, QI, Ft Hd Left, Em-None, Ft Dir Left, B-2

Probabilistic learning for English

Goal: Converge on English values after learning period is over

Learning Period Length: 1,666,667 words (based on estimates of words heard in a 6 month period, using Akhtar et al. (2004)).

QS, QSVCH, Em-Some, Em-Right, Ft Dir Right, Bounded, Bounded-2, Bounded-Syllabic, Ft Hd Left

Model	Success rate (1000 runs)
NParLearner, γ = .001, .0025, .01, .025	0.0%
BayesLearner	0.0%
NParLearner + Counting, γ = .001, .0025, .01, .025, c = 2, 5, 7, 10, 15, 20	0.033%
BayesLearner + Counting, c = 2, 5, 7, 10, 15, 20	0.0%

Acquirability results: parameters

- Four different implementations of reward/punishment tried (two Naïve Parameter Learner variants that use Linear reward-penalty schemes (Yang 2002) and two incremental Bayesian variants)

- Only one variant (one of the linear reward-penalty ones) was ever successful at converging on the adult English grammar, and then only once every 3000 runs! This seems like very poor performance from these cognitively inspired learners.

Problem with constrained learners?

- Maybe the problem is with the constrained learning algorithms: Are they identifying sub-optimal grammars for the data they encounter?
  - If so, ideal learners should find the optimal grammars that are most compatible with the English child-directed speech data

Premise: The adult English grammar is the grammar that best describes the systematic data of English, even if there are exceptions.

Implication: The adult English grammar is the grammar that is best able to generate the stress contours for the English data (most compatible).

- English grammar compatibility with data:
  - Generates contours matching 73.0% observable data tokens, where every instance of a word is counted (62.1% types, where frequency is factored out and a word is counted only once no matter how often it occurs)
  - Note: not expected to be at 100% because of irregularities in English data

Problem for any parametric learner

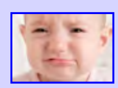
- Average compatibility of grammars selected by constrained learners:
  - 73.6% by tokens (63.3% by types)

  (Highest compatibility in hypothesis space: 76.5% by tokens, 70.3% by types)

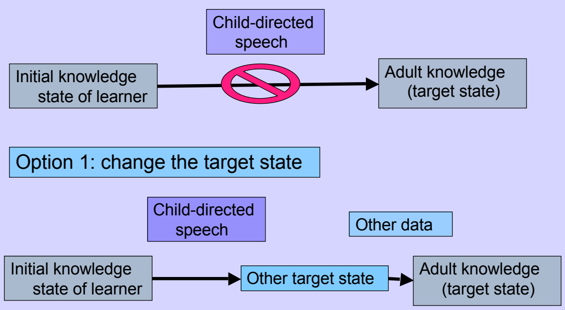
- The cognitively inspired learners *are* identifying the more optimal grammars for this data set - it's just that these grammars don't happen to be the adult English grammar!
  - Learnability Implication: The problem isn't because these learners are constrained. Unconstrained learners would have the same problem.
  - English grammar compared to other 155 grammars
    - Ranked 52nd by tokens, 56th by types
    - English grammar is barely in the top third - unsurprising that probabilistic learners rarely select this grammar, given the child-directed speech data!

Problem for any parametric learner

- Parametric child learner has a learnability problem: can't get to adult target state given the data available to children



Getting out of the learnability problem: 2 options



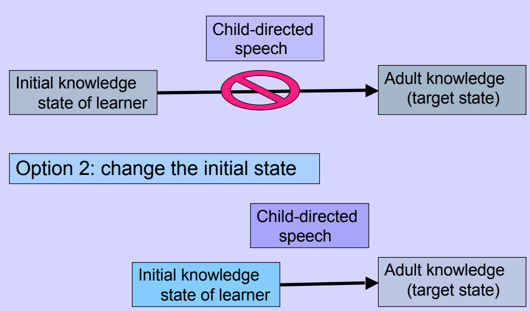
A different target state

- Maybe young children don't acquire the adult English grammar until later, after they are exposed to more word types and realize the connection between stress contour and the English morphological system (connection to English morphological system: Chomsky & Halle 1968, Kiparsky 1979, Hayes 1982)
  - Brown 1973: morphological inflections not used regularly till 36 months

    Prediction: Children initially select non-English grammars, given these data. If so, we should be able to use experimental methods to observe them using non-English grammars for an extended period of time.

  - Kehoe 1998: elicitation task with English 34-month-olds used items that were compatible with the grammars modeled learners often chose here

Getting out of the learnability problem: 2 options



A different (enriched) initial state

- Maybe young children have additional boosts
  - Pearl (2008) explores the effects of a bias to only learn from data perceived as unambiguous for a parametric learner, and finds that the learners with this knowledge are successful if parameters are set in certain orders.
  - Required knowledge at the initial state:
    - importance of unambiguous data (and a method for identifying these data for each parameter value)
    - parameter-setting order constraints (and potentially a method for deriving these constraints)

Unambiguous data bias

Why learning from unambiguous data works: The unambiguous data favor the English grammar, so English becomes the optimal grammar.

However, they make up a small percentage of the available data (never more than 5%) so their effect can be washed away in the wake of ambiguous data if the ambiguous data are learned from as well and the parameters are not learned in an appropriate order.

Bigger picture: Testing proposals of knowledge representation

- Began by exploring cognitively plausible learners to test theories about knowledge representation (argument from acquisition)
- When they failed at the acquisition task, we asked what the cause of the failure was - due to learners being constrained or due to something about the language acquisition computation?
- Led us to examine learnability considerations, given the data
  - Highlighted learnability issues for probabilistic learners seeking optimal solutions given child-directed speech data

A useful framework: what comes next

- Change knowledge representation
  - Theoretical + computational investigations: perhaps different parameters or constraints make the adult English grammar more acquirable from child-directed speech
  - Different theoretical proposals can be motivated and tested via computational methods

A useful framework: what comes next

- Change premise about trajectory of children's acquisition
  - Experimental investigations: exploring English children's initial knowledge states before they have knowledge of morphology and adult lexicon items
  - This then informs future computational investigations and thus any arguments from acquisition for a given theoretical proposal of knowledge representation

About that target state…

Analysis of adult-directed conversational speech

CALLFRIEND corpus (Canavan & Zipperlen 1996), North American English portion: recorded telephone conversations between adults

- 82,487 word tokens, 4,417 word types

Parametric English grammar (somewhat better but not the best):

- 63.7% token compatibility, 52.1% type compatibility
- ranked 34th by tokens, 36th by types
- Interesting: Best grammar in hypothesis space differs only by one parameter value (QI instead of English's QS): 66.6% token compatibility, 56.3% type compatibility

Parametric English grammar is not the best for adult conversational speech either

Potential explanation: linguists use items that appear infrequently in conversations when making their theories, under the assumption that these items are part of the adult knowledge state

Worth testing experimentally: the English adult knowledge state (do adults make the generalizations that linguists think they do, or are some of the crucial items exceptions that adults do not include in their generative system?)

A useful framework: what comes next

- Change learner's initial knowledge state
  - Computational investigations: strategies learners can use to solve acquisition problem as currently defined
  - Describe the required initial knowledge state to make acquisition possible for learners using specific knowledge representations, thereby creating a way to explicitly compare different knowledge representations
  - Knowledge representations requiring a less enriched initial state may be more desirable