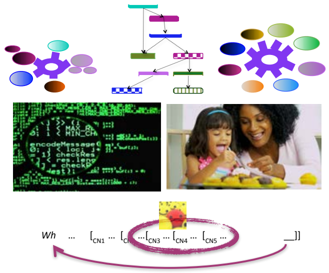
Evaluation, use, and refinement of knowledge representations through acquisition modeling

Lisa Pearl
University of California, Irvine

Computation of Language Laboratory
UC Irvine

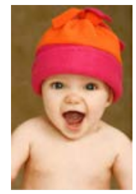


February 21, 2015:
Learning in Generative Grammar: 50 years since the evaluation metric
GALANA: University of Maryland, College Park

Premise

The knowledge representation provided by Universal Grammar is what makes acquisition happen so fast and so well.

Knowledge representation ↑ **Universal Grammar** → Language-specific grammar ↑ **Acquisition**

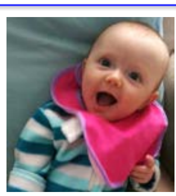


Why?

Premise

The knowledge representation provided by Universal Grammar is what makes acquisition happen so fast and so well.

Knowledge representation ↑ **Universal Grammar** = What hypotheses are available and how are they defined?



Grammar = ...
...set of parameters

...set of violable constraints

...rules over phrasal nodes

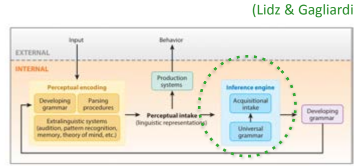
Premise

The knowledge representation provided by Universal Grammar is what makes acquisition happen so fast and so well.

Knowledge representation ↑ **Universal Grammar** = What hypotheses are available and how are they defined?

↓

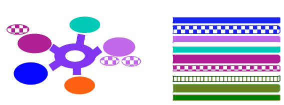
What part of the encoded input is relevant = **acquisitional intake**
(Lidz & Gagliardi 2015)



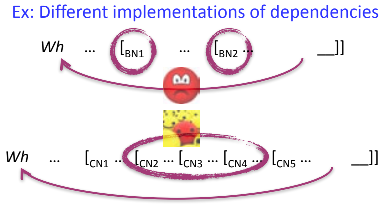
Competing theories

So how do we choose when we have multiple theories about how knowledge is represented?

Ex: Parameters vs. violable constraints



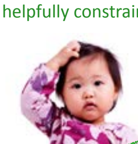
Ex: Different implementations of dependencies



Competing theories

One answer: Use each one for acquisition. Does that knowledge representation make acquisition possible from the available data?

Is the hypothesis space helpfully constrained by the representation?



Well...how do we tell?

Is the acquisitional intake defined by the hypothesis space sufficient to get the job done?

Well...how do we tell?

A helpful tool: Computational modeling

We can computationally model a learner who incorporates the assumptions of a representation, set that learner up in a cognitively plausible learning scenario, and see if acquisition succeeds.

Is the acquisitional intake defined by the hypothesis space sufficient to get the job done?

Well…how do we tell?

A helpful tool: Computational modeling

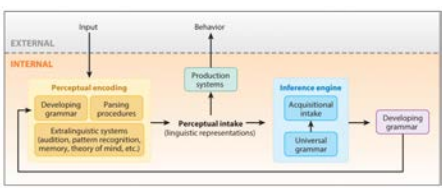
We can computationally model a learner who incorporates the assumptions of a representation, set that learner up in a cognitively plausible learning scenario, and see if acquisition succeeds.

What we learn from computational modeling:

- Which representations allow acquisition to succeed
- What needs to be true about the learning scenario for a learner to succeed using those representations

The goal

This computational modeling feedback helps us refine our theories about both the knowledge representation and the acquisition process that uses that representation.

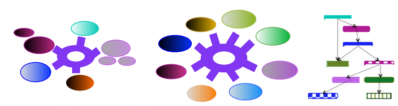


(Lidz & Gagliardi 2015)

Today's goal: Computational acquisition modeling

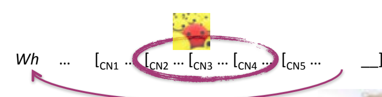
Case studies

Metrical phonology



OCtopus

Syntax

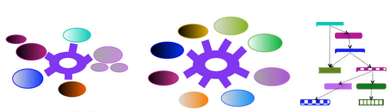


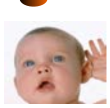
What did you see?

Today's goal: Computational acquisition modeling

Case studies

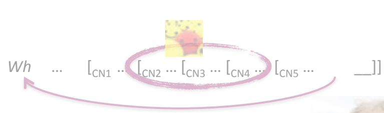
Metrical phonology





OCtopus

Syntax



What did you see?

Metrical phonology: Target knowledge

Account for word-level stress patterns

Observable data: stress contour

OCtopus

Underlying representation?

Points of agreement:

Use metrical feet:
Units ≥ syllables
but (often) smaller than words

(…) (………)

Look only at syllable rimes

VC V VC
ɑk ə ʊs

Divide word into syllables

ɑk tə pʊs
oc to pus

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Target knowledge

Account for word-level stress patterns

Observable data: stress contour   OCtopus

Underlying representation?

Points of cross-linguistic variation:

?

- How to classify syllables
- What metrical feet are allowed
- How stress interacts with metrical feet

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

VC V VC
ɑk ə ʊs
ɑk tə pʊs
oc to pus

Pearl, Ho, & Detrano 2014, under rev.

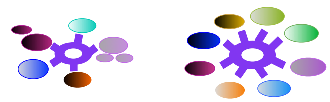
Metrical phonology: Target knowledge

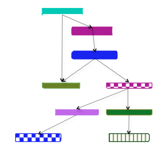
Account for word-level stress patterns

Observable data: stress contour   OCtopus

Points of disagreement: Underlying grammar = ….?

Parameters with values set   Ranked violable constraints





Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Parametric systems

Correct grammar builds compatible contour

OCtopus

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Parametric systems

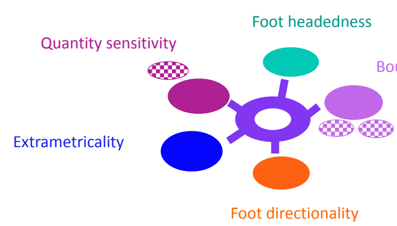
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Grammar = Set of parameter & sub-parameter values

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Parametric systems

Correct grammar builds compatible contour

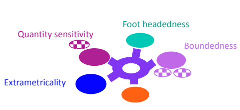
OCtopus

This grammar, comprised of particular parameter values, generates the correct stress contour.

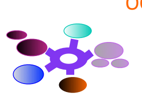
HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



(H L) H
OC to pus



Parameter values used:
QS-VC-H, Em-Rt, FtDir-Rt, B-2-Syl, FtHd-Left

…which are the values of the English grammar.

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Parametric systems

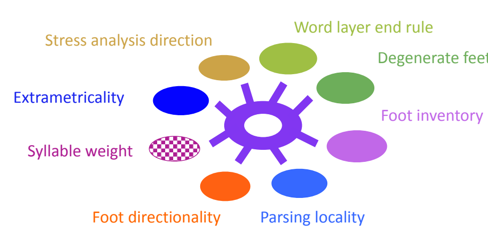
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

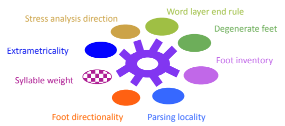
Parametric systems

Correct grammar builds compatible contour

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



OCtopus

This grammar, comprised of particular parameter values, generates an incorrect stress contour.

(H)(L L)
OC TÓ pus

Parameter values used:
Bot, Em-RtCons, VC-H, FtDir-Rt, PL-Strong, MorTro, DF-Strong, WLER-Rt

...which are the values of the English grammar.

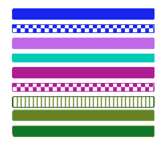
Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

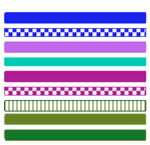
Best candidate for the correct grammar has a compatible contour

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

OCtopus

Premise: Many different candidates for a word's stress representation and contour are generated and then ranked according to which constraints are violated. Violating higher-ranked constraints is worse than violating lower-ranked constraints.



	C1	C2	C3	C4
(OC to) pus			*	*
oc (TO pus)	*		*	
(oc TO) pus		*	*	

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

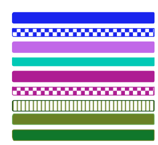
Best candidate for the correct grammar has a compatible contour

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

OCtopus



Grammar = ranked ordering of all constraints

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

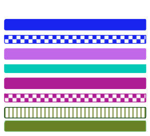
OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

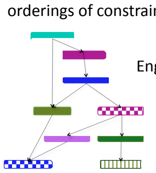
9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

OCtopus

Official grammars for languages are often described as partial orderings of constraints.





English grammar

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

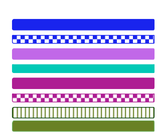
OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

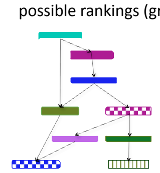
9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

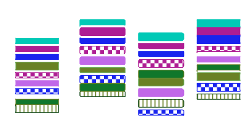
OCtopus

This means the "grammar" for a language is often a set of the possible rankings (grammars) that obey those orderings.





Ex: The English "grammar" is compatible with 26 rankings.



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

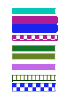
9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Sample candidates

A sample grammar that is a version of the English "grammar":



(OC to) (PUS) | (OC to) pus

(oc TO) (PUS) | oc (TO pus)

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Only one candidate left, and it has a compatible contour.

Sample candidates

A sample grammar that is a version of the English "grammar":

(OC to) (PUS) | (OC to) pus

(oc TO) (PUS) | oc (TO pus)

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

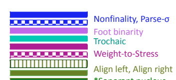
OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

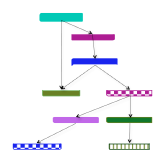
9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress



A sample grammar that is a version of the English "grammar":



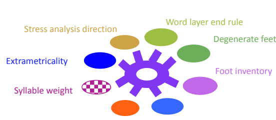
English "grammar"

Pearl, Ho, & Detrano 2014, under rev.

Knowledge representation comparison

HV: 5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Hayes: 8 parameters

Hypothesis space: 768 grammars

OT: 9 violable constraints

Hypothesis space: 362,880 grammars

Pearl, Ho, & Detrano 2014, under rev.

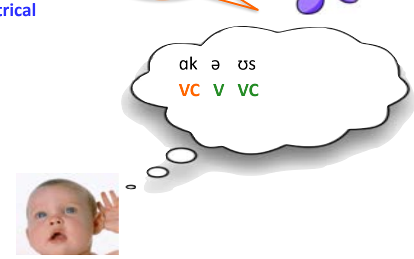
Metrical phonology: Acquisitional intake

Acquisition goal: Identify the grammar that can account for the word-level stress patterns in the language

Observable data: stress contour

OCtopus

All representations: use metrical feet based on syllable rimes



Pearl, Ho, & Detrano 2014, under rev.

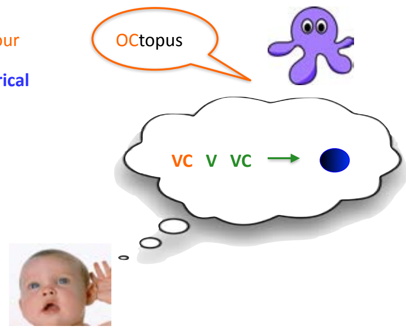
Metrical phonology: Acquisitional intake

Acquisition goal: Identify the grammar that can account for the word-level stress patterns in the language

Observable data: stress contour

All representations: use metrical feet based on syllable rimes

Parametric inference:
Does this set any values?



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Acquisitional intake

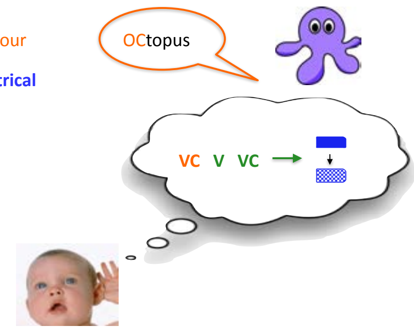
Acquisition goal: Identify the grammar that can account for the word-level stress patterns in the language

Observable data: stress contour

All representations: use metrical feet based on syllable rimes

Parametric inference:
Does this set any values?

OT inference:
Does this implicate any constraint rankings?

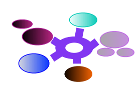


Pearl, Ho, & Detrano 2014, under rev.

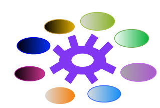
Computational acquisition evaluation: English

Pearl, Ho, & Detrano 2014, under rev.

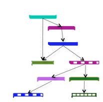
English grammar



HV: 5 parameters & 3 sub-parameters
Hypothesis space: 156 grammars



Hayes: 8 parameters
Hypothesis space: 768 grammars



OT: 9 violable constraints
Hypothesis space: 362,880 grammars
(English = 26 grammars)

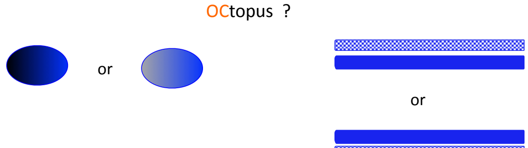
Note: These values/rankings are derived from stress patterns for English words in the adult lexicon

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Non-trivial because there are many data that are ambiguous for which parameter value or constraint ranking they implicate

OCtopus ?



This is generally a problem for acquisition (poverty of the stimulus = the data are compatible with many hypotheses).

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Non-trivial because there are many irregularities. This is less common for acquisition – usually there aren't a lot of exceptions to the system being acquired.

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Non-trivial because there are many irregularities. This is less common for acquisition – usually there aren't a lot of exceptions to the system being acquired.

Some causes of irregularity:

Interactions with morphology (Chomsky & Halle 1968, Hayes 1982, Kiparsky 1979)

Example: Adding productive morphology doesn't change the stress pattern, even though all grammars base their stress patterns on the syllables present in the word.

EARly	PREtty	senSAtion
EARlier	PREttiest	senSAtional
		senSAtionally

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Non-trivial because there are many irregularities. This is less common for acquisition – usually there aren't a lot of exceptions to the system being acquired.

Some causes of irregularity:

Interactions with grammatical category (Hammond 1999, Hayes 1982, Cassidy & Kelly 2001, Christiansen & Monaghan 2006)

Stress contours may be different across grammatical categories, even though the syllabic word form doesn't change.

NOUNS	VERBS	Syllabic word form
CONduct	conDUCT	VC VCC
DEsert	deSERT	V VCC
SUspect	suSPECT	V VCC

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

These irregularities can cause multiple stress contours to be associated with a syllabic word form. This is problematic for the grammars in these knowledge representations…

Syllabic word form: V VV

KI tty	a WAY	UH OH
V vv	v VV	V VV

Pearl, Ho, & Detrano 2014, under rev.

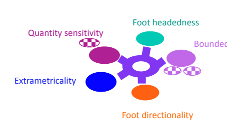
Learning English metrical phonology: Non-trivial

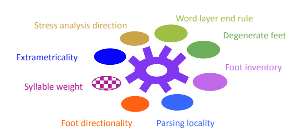
These irregularities can cause multiple stress contours to be associated with a syllabic word form. This is problematic for the grammars in these knowledge representations, since a grammar can only generate a single stress contour per syllabic word form…

Syllabic word form: V VV

Generate one of these…

KI tty	a WAY	UH OH
V vv	v VV	V VV





Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

These irregularities can cause multiple stress contours to be associated with a syllabic word form. This is problematic for the grammars in these knowledge representations, since a grammar can only generate a single stress contour per syllabic word form or select a single stressed syllabic word form as the best candidate.

Syllabic word form: V VV

Select one of these…

KI tty	a WAY	UH OH
V vv	v VV	V VV

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Upshot of multiple stress contours: No one grammar can account for all the stressed words in the input.

But how big of a problem is this in English child-directed speech?

Syllabic word form: V VV

KI tty	a WAY	UH OH
V vv	v VV	V VV

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Non-trivial

Analysis of Brent corpus (CHILDES database): 4780 word types (99,968 tokens) of American English speech directed at children between the ages of 6 and 12 months

Syllabic word form: V VV

KI tty	a WAY	UH OH
V vv	v VV	V VV

Multiple stress contours

HV: 73 of 123 syllabic word forms
Hayes: 86 of 149 syllabic word forms
OT: 166 of 452 syllabic word forms

This occurs a lot!

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Target state update

Acquisition success: Identify the grammar that can account for ^ the word-level stress patterns in the language

a good portion of

Is this reasonable?

Probably.

A grammar is useful because it provides a compact representation of some aspect of the data. Even if it doesn't cover all the data, covering some is helpful.

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Acquisition evaluation

How easily does a knowledge representation allow children to learn their specific language's grammar, when given realistic data?

Learnability analysis provides a quantitative way to compare competing knowledge representations (Pearl 2011, Legate & Yang 2012)

Working premise: Rational learners

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Acquisition evaluation

How easily does a knowledge representation allow children to learn their specific language's grammar, when given realistic data?

Learnability analysis provides a quantitative way to compare competing knowledge representations (Pearl 2011, Legate & Yang 2012)

Working premise: Rational learners

A learner trying to learn which grammar is the right one for the language will choose the grammar perceived to be the best.

Pearl, Ho, & Detrano 2014, under rev.

Learning English metrical phonology: Acquisition evaluation

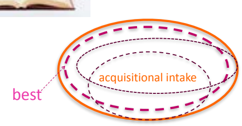
How easily does a knowledge representation allow children to learn their specific language's grammar, when given realistic data?

Learnability analysis provides a quantitative way to compare competing knowledge representations (Pearl 2011, Legate & Yang 2012)

Working premise: Rational learners

A learner trying to learn which grammar is the right one for the language will choose the grammar perceived to be the best.

able to account for the most data in the acquisitional intake = most useful to have



Pearl, Ho, & Detrano 2014, under rev.

Quantifying learnability

Once we define the acquisitional intake, we can then ask which grammar in the hypothesis space defined by the knowledge representation is best, assuming a rational learner that will choose the grammar compatible with the most data.

Pearl, Ho, & Detrano 2014, under rev.

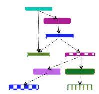
Quantifying learnability

Once we define the acquisitional intake, we can then ask which grammar in the hypothesis space defined by the knowledge representation is best, assuming a rational learner that will choose the grammar compatible with the most data.

Compatibility with a data point: A grammar is compatible with a data point if the grammar can account for that data point.

Here: Matching stress contour.

Individual grammar evaluation



Intuition: More compatibility is better.
A grammar that can account for 70% of the data is better than a grammar that can only account for 55% of the data.

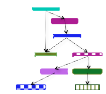
Pearl, Ho, & Detrano 2014, under rev.

Quantifying learnability

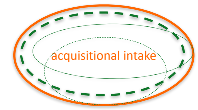
Once we define the acquisitional intake, we can then ask which grammar in the hypothesis space defined by the knowledge representation is best, assuming a rational learner that will choose the grammar compatible with the most data.

Raw compatibility for a grammar: The amount of data that grammar can account for.

Individual grammar evaluation



Example: A grammar that can account for 70% of the data has a raw compatibility of 0.70.



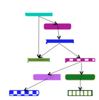
Pearl, Ho, & Detrano 2014, under rev.

Quantifying learnability

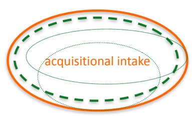
Once we define the acquisitional intake, we can then ask which grammar in the hypothesis space defined by the knowledge representation is best, assuming a rational learner that will choose the grammar compatible with the most data.

Relative compatibility for a grammar: The proportion of other grammars that this grammar is better than, based on raw compatibility.

Individual grammar evaluation



Example: The best grammar in the knowledge representation has ~1.00 relative compatibility. This is the one that's easiest to learn, given the data.



Pearl, Ho, & Detrano 2014, under rev.

Quantifying learnability

Once we define the acquisitional intake, we can then ask which grammar in the hypothesis space defined by the knowledge representation is best, assuming a rational learner that will choose the grammar compatible with the most data.

Learnability potential for a knowledge representation: The amount of data the best grammar is compatible with. This is how much of the data that knowledge representation is capable of accounting for with any of its grammars.

Knowledge representation evaluation

Example: If the best grammar can account for 70% of the data, this knowledge representation has a learnability potential of 0.70.

Pearl, Ho, & Detrano 2014, under rev.

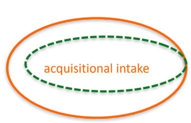
English learnability: Knowledge representations

So what's the best any grammar in a given knowledge representation actually does, given realistic child-directed data?

Learnability potential = proportion of data the best grammar (relative compatibility ≈ 1.00) can account for

Raw compatibility of best grammar
HV: 0.668 types
Hayes: 0.683 types
OT: 0.657 types

Around 2/3 of the word types



Pearl, Ho, & Detrano 2014, under rev.

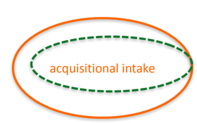
English learnability: Knowledge representations

Implication:

The best grammar in any of these knowledge representations is pretty useful to have. It allows a learner to account for a good proportion of the input, even if there's a significant chunk that can't be accounted for.

Raw compatibility of best grammar
HV: 0.668 types
Hayes: 0.683 types
OT: 0.657 types

Around 2/3 of the word types



Pearl, Ho, & Detrano 2014, under rev.

English learnability: Knowledge representations

Implication:

The best grammar in any of these knowledge representations is pretty useful to have. It allows a learner to account for a good proportion of the input, even if there's a significant chunk that can't be accounted for.

But…is that really the best they can do?

Pearl, Ho, & Detrano 2014, under rev.

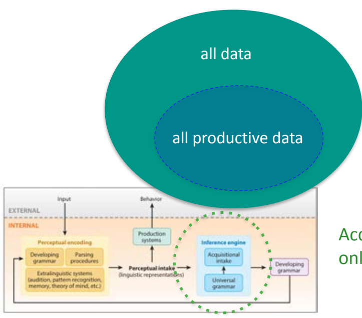
Data filters

Previous working assumption: The learner will try to learn a grammar that can account for all the data encountered.

Pearl, Ho, & Detrano 2014, under rev.

Data filters

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).



(Lidz & Gagliardi 2015)

Acquisitional intake = only productive data

Pearl, Ho, & Detrano 2014, under rev.

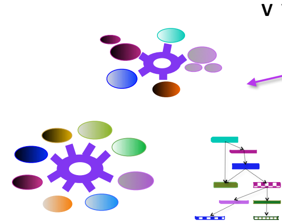
Productive data filter for metrical phonology

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).

Principled way to implement this = Tolerance Principle

Syllabic word form: V VV

✓ COO kie, DA ddy, FU nny



✗ a WAY, be LOW, to DAY

Productive data filter

Pearl, Ho, & Detrano 2014, under rev.

English learnability: Knowledge representations

So what's the best any grammar in a given knowledge representation actually does, given realistic child-directed data and a productive data filter?

Learnability potential = proportion of data the best grammar (relative compatibility ≈ 1.00) can account for

Raw compatibility of best grammar

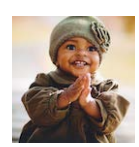
HV: 0.949 productive types
Hayes: 0.933 productive types
OT: 0.843 productive types

84-95% of the productive word types

Pearl, Ho, & Detrano 2014, under rev.

English learnability: Knowledge representations

This looks even better! Though the parametric representations are a little ahead, all representations can generate a grammar that accounts for a very large proportion of the productive data. Quite useful!



Raw compatibility of best grammar

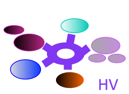
HV: 0.949 productive types
Hayes: 0.933 productive types
OT: 0.843 productive types

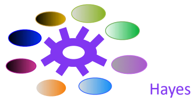
84-95% of the productive word types

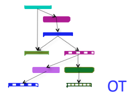
Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What about the "English" grammar in each knowledge representation?







Working assumption for acquisition:
The "English" grammar should be the best grammar in a representation (relative compatibility ≈ 1.00) for the data of English.

Is it?

Pearl, Ho, & Detrano 2014, under rev.

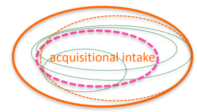
English learnability: English grammars

How does the (best) English grammar compare to the other grammars defined by the knowledge representation, when looking at all the data?

Relative compatibility of the English grammar = proportion of grammars in the hypothesis space the (best) English grammar is better than

Relative compatibility of English grammar
HV: 0.673 out of 156 grammars
Hayes: 0.676 out of 768 grammars
OT: 0.817 out of 362,880 grammars

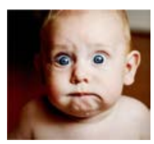
Better than many…but many are still better
HV: 51 are better
Hayes: 249 are better
OT: 66,407 are better



Pearl, Ho, & Detrano 2014, under rev.

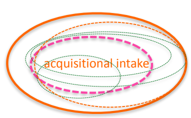
English learnability: English grammars

Implication:
There are many other grammars in the hypothesis space that are more compatible with the data. It would be easier to pick one of these other more compatible grammars.



Relative compatibility of English grammar
HV: 0.673 out of 156 grammars
Hayes: 0.676 out of 768 grammars
OT: 0.817 out of 362,880 grammars

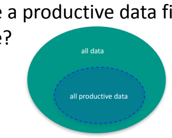
Better than many…but many are still better
HV: 51 are better
Hayes: 249 are better
OT: 66,407 are better



Pearl, Ho, & Detrano 2014, under rev.

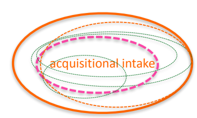
English learnability: English grammars

What about if children have a productive data filter on their acquisitional intake?



Relative compatibility of English grammar
HV: 0.622 out of 156 grammars
Hayes: 0.680 out of 768 grammars
OT: 0.798 out of 362,880 grammars

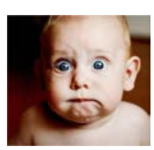
Little if any improvement…and sometimes worse:
HV: 59 are better
Hayes: 246 are better
OT: 73,302 are better



Pearl, Ho, & Detrano 2014, under rev.

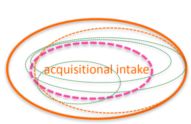
English learnability: English grammars

Implication:
There are still many other grammars in the hypothesis space that are more compatible with the data. It would be easier to pick one of these other more compatible grammars.



Relative compatibility of English grammar
HV: 0.622 out of 156 grammars
Hayes: 0.680 out of 768 grammars
OT: 0.798 out of 362,880 grammars

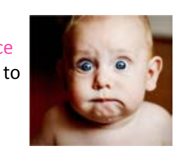
Little if any improvement…and sometimes worse:
HV: 59 are better
Hayes: 246 are better
OT: 73,302 are better



Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

Implication:
There are still many other grammars in the hypothesis space that are more compatible ~~with the data.~~ It would be easier to pick one of these other more compatible grammars.



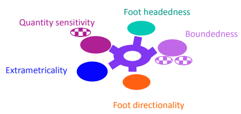
So what do these other more compatible grammars look like anyway?
What values/constraint rankings do they use?
Can the current English grammar definitions be adjusted?

Pearl, Ho, & Detrano 2014, under rev.

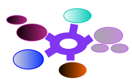
English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: HV



English grammar

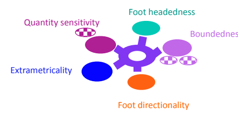


Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: HV



It turns out that many high compatibility grammars use a different Quantity Sensitivity value: Quantity Insensitive (QI), rather than Quantity Sensitive (QS).

This allows them to handle words like *bellybutton*, which have an unstressed internal heavy syllable.

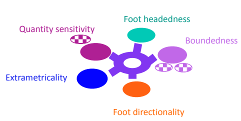
BE lly BU tton

Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

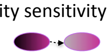
What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: HV



It turns out that many high compatibility grammars use a different Quantity Sensitivity value: Quantity Insensitive (QI), rather than Quantity Sensitive (QS).

So what happens if we swap the English definition's quantity sensitivity value?

QS —> QI 

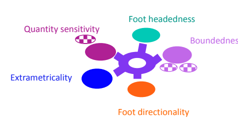
Relative compatibility over all data = 0.94!

Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

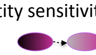
What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

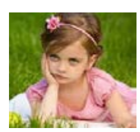
Parametric: HV



It turns out that many high compatibility grammars use a different Quantity Sensitivity value: Quantity Insensitive (QI), rather than Quantity Sensitive (QS).

So what happens if we swap the English definition's quantity sensitivity value?

QS —> QI 



Relative compatibility over all data = 0.94!

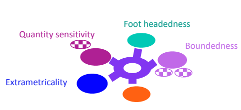
But relative compatibility over productive data = 0.71…

Pearl, Ho, & Detrano 2014, under rev.

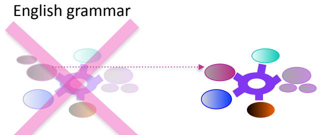
English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: HV



English grammar



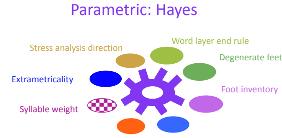
Upshot: For the HV knowledge representation, the learning problem could be ameliorated by simply switching one parameter value as long as children aren't using a productive data filter.

Pearl, Ho, & Detrano 2014, under rev.

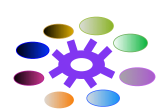
English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: Hayes



English grammar

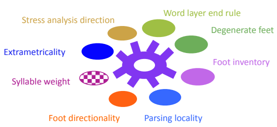


Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: Hayes



It turns out that many high compatibility grammars use a different Foot Inventory value: Syllabic Trochees (Syl-Tro) rather than Moraic Trochees (Mor-Tro).

This allows them to handle words like *baby* and *kitty*, which have a final unstressed heavy syllable.

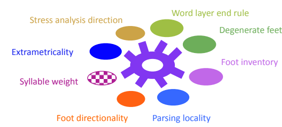
BA by
KI tty

Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

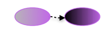
Parametric: Hayes



It turns out that many high compatibility grammars use a different Foot Inventory value: Syllabic Trochees (Syl-Tro) rather than Moraic Trochees (Mor-Tro).

So what happens if we swap the English definition's foot inventory value?

Mor-Tro —> Syl-Tro



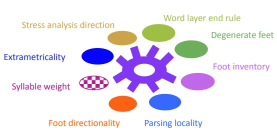
Relative compatibility over all data = 0.91

Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

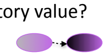
Parametric: Hayes



It turns out that many high compatibility grammars use a different Foot Inventory value: Syllabic Trochees (Syl-Tro) rather than Moraic Trochees (Mor-Tro).

So what happens if we swap the English definition's foot inventory value?

Mor-Tro —> Syl-Tro



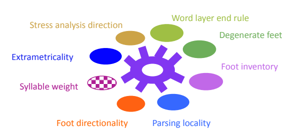
Relative compatibility over all data = 0.91
And relative compatibility over productive data = 0.96!

Pearl, Ho, & Detrano 2014, under rev.

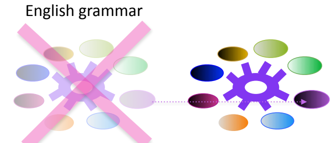
English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Parametric: Hayes



English grammar



Upshot: For the Hayes knowledge representation, the learning problem could be ameliorated by simply switching one parameter value especially if children are using a productive data filter.

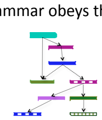
Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Constraint-based: OT

English grammar obeys this ordering

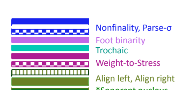


Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Constraint-based: OT



It turns out that **all** high compatibility grammars use a different ordering of Non-Finality (Non-Fin) and Weight-to-Stress VV (WSP-VV):
Non-Fin is ranked higher than WSP-VV

This allows them to handle words like *baby* and *kitty*, which have a final unstressed VV syllable.

BA by
KI tty

Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Constraint-based: OT

It turns out that **all** high compatibility grammars use a different ordering of Non-Finality (Non-Fin) and Weight-to-Stress VV (WSP-VV): Non-Fin is ranked higher than WSP-VV

So what happens if we swap the English definition's ordering of these constraints?

Non-Fin >> WSP-VV

Relative compatibility over all data = 0.99!

Pearl, Ho, & Detrano 2014, under rev.

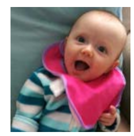
English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Constraint-based: OT

It turns out that **all** high compatibility grammars use a different ordering of Non-Finality (Non-Fin) and Weight-to-Stress VV (WSP-VV): Non-Fin is ranked higher than WSP-VV

So what happens if we swap the English definition's ordering of these constraints?



Non-Fin >> WSP-VV

Relative compatibility over all data = 0.99!

Relative compatibility over productive data = 0.93!

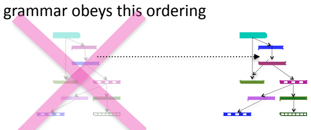
Pearl, Ho, & Detrano 2014, under rev.

English learnability: English grammars

What values/constraint rankings do the grammars use that are more compatible with the data than the official English grammar?

Constraint-based: OT

English grammar obeys this ordering



Upshot: For the OT knowledge representation, the learning problem could be ameliorated by simply switching one constraint ordering especially if children aren't using a productive data filter.

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology learnability: Take away

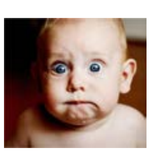
✓ All three knowledge representations are useful for acquisition: They can generate grammars that account for a large portion of realistic English child-directed speech data, especially if children are using a productive data filter.

Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology learnability: Take away

✓ All three knowledge representations are useful for acquisition especially if children are using a productive data filter.

However, the current English grammar definitions in each representation are not the grammars most easily learnable from the data.



Pearl, Ho, & Detrano 2014, under rev.

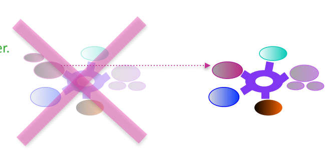
Metrical phonology learnability: Take away

✓ All three knowledge representations are useful for acquisition especially if children are using a productive data filter.

However, the current English grammar definitions in each representation are not the grammars most easily learnable from the data.

But each representation has a grammar that is very close to the current English grammar definition (change one parameter value or one constraint ordering) which *is* much more easily learnable.

Parametric: HV
Fine only if children aren't using a productive data filter.



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology learnability: Take away

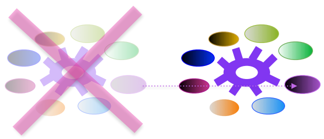
✓ All three knowledge representations are useful for acquisition especially if children are using a productive data filter.

However, the current English grammar definitions in each representation are not the grammars most easily learnable from the data.

But each representation has a grammar that is very close to the current English grammar definition (change one parameter value or one constraint ordering) which *is* much more easily learnable.

Parametric: HV
Fine only if children aren't using a productive data filter.

Parametric: Hayes
Better if children are using a productive data filter.



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology learnability: Take away

✓ All three knowledge representations are useful for acquisition especially if children are using a productive data filter.

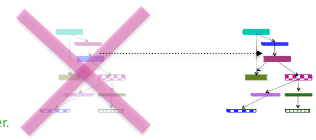
However, the current English grammar definitions in each representation are not the grammars most easily learnable from the data.

✓ But each representation has a grammar that is very close to the current English grammar definition (change one parameter value or one constraint ordering) which *is* much more easily learnable.

Parametric: HV
Fine only if children aren't using a productive data filter.

Parametric: Hayes
Better if children are using a productive data filter.

Constraint-based: OT
A little better if children aren't using a productive data filter.



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology learnability: Take away

Open questions:

Theory of acquisition:
Are children using a productive data filter? This affects how much data any representation can account for and which particular representation has an English-like grammar that is easily learnable from realistic child-directed English data.

Parametric: HV
Fine only if children aren't using a productive data filter.

Parametric: Hayes
Better if children are using a productive data filter.

Constraint-based: OT
A little better if children aren't using a productive data filter.

Pearl, Ho, & Detrano 2014, under rev.

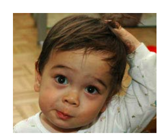
Metrical phonology learnability: Take away

Open questions:

Theory of representation for English:
Are these English-like grammars the ones children have? Are these the ones adults have? Can verify experimentally.

Are these English-like grammars (more?) compatible with adult-directed English data? Can verify computationally.

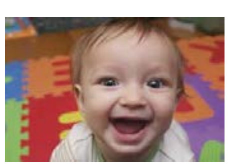
If so, this supports these grammars as the actual English grammar in each representation.



Pearl, Ho, & Detrano 2014, under rev.

Metrical phonology: Big picture

This approach allows us to evaluate metrical phonology representations by using them for acquisition. We can then refine our theories of acquisition and representation.



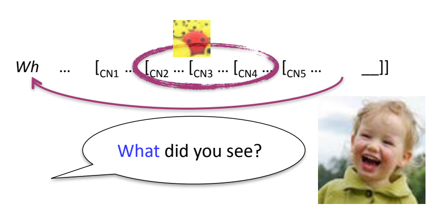
Pearl, Ho, & Detrano 2014, under rev.

Today's goal: Computational acquisition modeling

Case studies

Metrical phonology

Syntax



Syntax: Syntactic islands

- **Why**? Central to UG-based syntactic theories.

- **What**? Dependencies can exist between two non-adjacent items. They do not appear to be constrained by length (Chomsky 1965, Ross 1967), but rather by whether the dependency crosses certain structures (called "syntactic islands").

What does Jack think __?

What does Jack think that Lily said that Sarah heard that Jareth believed __?

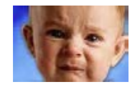
Pearl & Sprouse 2013a, 2013b, under review

Syntax: Syntactic islands

- **Why**? Central to UG-based syntactic theories.

- **What**? Dependencies can exist between two non-adjacent items. They do not appear to be constrained by length (Chomsky 1965, Ross 1967), but rather by whether the dependency crosses certain structures (called "syntactic islands").

<u>Some example islands</u>



Complex NP island:
*What did you make [the claim that Jack bought __]?
Subject island:
*What do you think [the joke about __] offended Jack?
Whether island:
*What do you wonder [whether Jack bought __]?
Adjunct island:
*What do you worry [if Jack buys __]?

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

What does Jack think __?

What does Jack think that Lily said that Sarah heard that Jareth believed __?

Complex NP island:
*What did you make [the claim that Jack bought __]?
Subject island:
*What do you think [the joke about __] offended Jack?
Whether island:
*What do you wonder [whether Jack bought __]?
Adjunct island:
*What do you worry [if Jack buys __]?

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Complex NP islands

Who __ claimed that Lily forgot the necklace?	matrix | non-island
What did the teacher claim that Lily forgot __?	embedded | non-island
Who __ made the claim that Lily forgot the necklace?	matrix | island
*What did the teacher make the claim that Lily forgot __?	embedded | island

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Subject islands

Who __ thinks the necklace is expensive?	matrix | non-island
What does Jack think __ is expensive?	embedded | non-island
Who __ thinks the necklace for Lily is expensive?	matrix | island
*Who does Jack think the necklace for __ is expensive?	embedded | island

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Whether islands

Who __ thinks that Jack stole the necklace?	matrix | non-island
What does the teacher think that Jack stole __ ?	embedded | non-island
Who __ wonders whether Jack stole the necklace?	matrix | island
*What does the teacher wonder whether Jack stole __ ?	embedded | island

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:
- length of dependency (matrix vs. embedded)
- presence of an island structure (non-island vs. island)

Adjunct islands

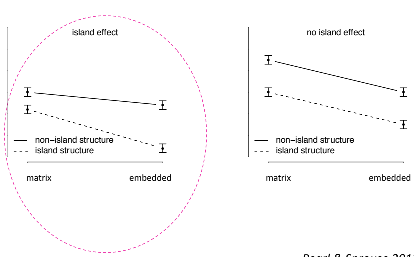
Who __ thinks that Lily forgot the necklace?	matrix | non-island
What does the teacher think that Lily forgot __ ?	embedded | non-island
Who __ worries if Lily forgot the necklace?	matrix | island
*What does the teacher worry if Lily forgot __ ?	embedded | island

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Syntactic island = superadditive interaction of the two factors (additional unacceptability that arises when the two factors are combined, above and beyond the independent contribution of each factor).



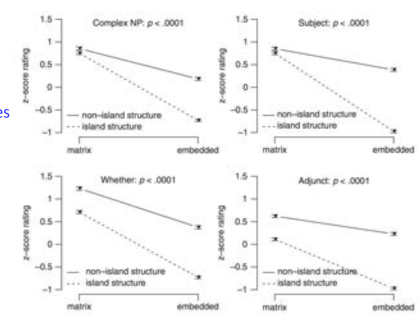
Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior
(from Sprouse et al. 2012 data on the four island types, with 173 subjects)

Superadditivity present for all islands tested =
Knowledge that dependencies cannot cross these island structures is part of adult knowledge about syntactic islands

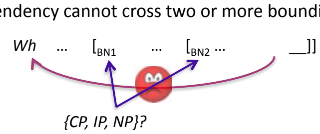
Importance for acquisition: This is one kind of target behavior that we'd like a learner to produce.

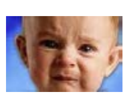


Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



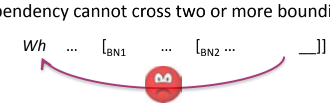


Bounding nodes are language-specific
(CP, IP, and/or NP – must learn which ones are relevant for language)

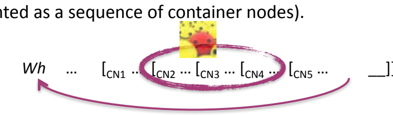
Pearl & Sprouse 2013a, 2013b, under review

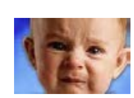
Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b)
(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).





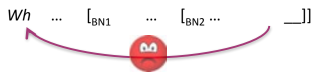
Container node: phrase structure node that contains dependency

[CP *What do* [IP *you* [VP *like* __ [PP *in this picture?*]]]]

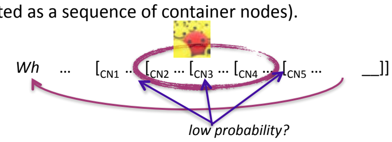
Pearl & Sprouse 2013a, 2013b, under review

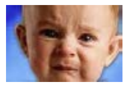
Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b)
(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



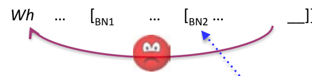


Low probability regions are language-specific
(defined by sequences of container nodes that must be learned)

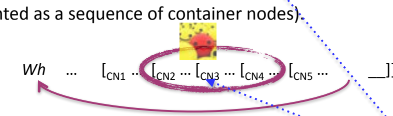
Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b)
(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

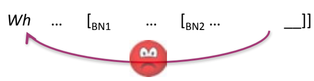


In common: Both rely on local structure anomalies (at some level)

Pearl & Sprouse 2013a, 2013b, under review

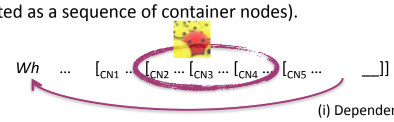
Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



(i) Dependencies defined over bounding nodes — track those
(ii) Bounding node = ?
(iii) 2+ bounding nodes =

Subjacency-ish (Pearl & Sprouse 2013a, 2013b)
(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



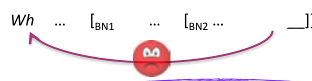
(i) Dependencies defined over container node structure — track that already
(ii) Container node = ?
(iii) low probability =

Different: Amount of language-specific knowledge built in just for islands

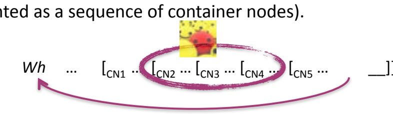
Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)
(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b)
(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

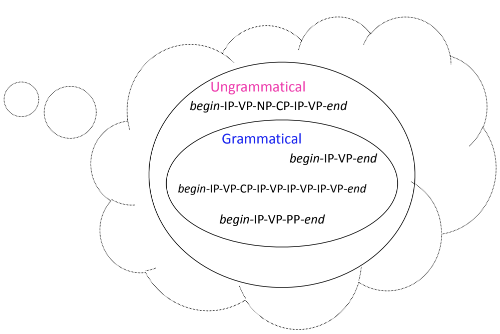


Focus on evaluating this one today

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Hypothesis space

Children's hypotheses are about what container node sequences are grammatical for dependencies in the language.



Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Target knowledge

Can the grammatical dependencies be distinguished from the ungrammatical ones?

Sprouse et al. (2012) stimuli:

Complex NP islands		**Subject islands**
begin-IP-end	matrix | non-island	*begin-IP-end*
begin-IP-VP-CP-IP-VP-end	embedded | non-island	*begin-IP-VP-CP-IP-end*
begin-IP-end	matrix | island	*begin-IP-end*
**begin-IP-VP-NP-CP-IP-VP-end*	embedded | island	**begin-IP-VP-CP-IP-NP-PP-end*

✓ All the ungrammatical dependencies are distinct from all the grammatical dependencies for these syntactic islands.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Target knowledge

Can the grammatical dependencies be distinguished from the ungrammatical ones?

Sprouse et al. (2012) stimuli:

Whether islands		Adjunct islands
begin-IP-*end*	matrix | non-island	*begin*-IP-*end*
begin-IP-VP-CP-IP-VP-*end*	embedded | non-island	*begin*-IP-VP-CP-IP-VP-*end*
begin-IP-*end*	matrix | island	*begin*-IP-*end*
**begin*-IP-VP-CP-IP-VP-*end*	embedded | island	**begin*-IP-VP-CP-IP-VP-*end*

Uh oh - the ungrammatical dependencies look identical to some of the grammatical dependencies for these syntactic islands.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Dependency representation

One solution:
Have CP container nodes be more specified for the learner:
Use the lexical head to subcategorize the CP container node.

CP$_{null}$, CP$_{that}$, CP$_{whether}$, CP$_{if}$, etc.

The learner can then distinguish between these structures:

IP-VP-CP$_{null/that}$-IP-VP
IP-VP-CP$_{whether/if}$-IP-VP

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Target knowledge

Can the grammatical dependencies be distinguished from the ungrammatical ones?

Sprouse et al. (2012) stimuli:

Complex NP islands		Subject islands
begin-IP-*end*	matrix | non-island	*begin*-IP-*end*
begin-IP-VP-CP$_{that}$-IP-VP-*end*	embedded | non-island	*begin*-IP-VP-CP$_{null}$-IP-*end*
begin-IP-*end*	matrix | island	*begin*-IP-*end*
**begin*-IP-VP-NP-CP$_{that}$-IP-VP-*end*	embedded | island	**begin*-IP-VP-CP$_{null}$-IP-NP-PP-*end*

✓ All the ungrammatical dependencies are still distinct from all the grammatical dependencies for these syntactic islands.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Target knowledge

Can the grammatical dependencies be distinguished from the ungrammatical ones?

Sprouse et al. (2012) stimuli:

Whether islands		Adjunct islands
begin-IP-*end*	matrix | non-island	*begin*-IP-*end*
begin-IP-VP-CP$_{that}$-IP-VP-*end*	embedded | non-island	*begin*-IP-VP-CP$_{that}$-IP-VP-*end*
begin-IP-*end*	matrix | island	*begin*-IP-*end*
**begin*-IP-VP-CP$_{whether}$-IP-VP-*end*	embedded | island	**begin*-IP-VP-CP$_{if}$-IP-VP-*end*

✓ Now the ungrammatical dependencies are distinct from all the grammatical dependencies for these syntactic islands, too.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Acquisitional intake

Children must learn which local pieces of structure are low probability for a *wh*-dependency. They learn this from the *wh*-dependencies in their intake, which are defined over the container nodes of the *wh*-dependency.

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ like __]]]?
IP VP

Encoding of dependency: *begin*-IP-VP-*end*

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?
IP VP CP$_{null}$ IP VP PP

Encoding of dependency: *begin*-IP-VP-CP$_{null}$-IP-VP-PP-*end*

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Realistic acquisitional intake

Child-directed speech (Brown-Adam, Brown-Eve, Suppes, Valian) from CHILDES: 101,838 utterances containing 20,923 *wh*-dependencies

What kind of dependencies are present?

76.7%	*begin*-IP-VP-*end*	*What did you see __?*
12.8%	*begin*-IP-*end*	*What __ happened?*
5.6%	*begin*-IP-VP-IP-VP-*end*	*What did she want to do __?*
2.5%	*begin*-IP-VP-PP-*end*	*What did she read from __?*
1.1%	*begin*-IP-VP-CP$_{null}$-IP-VP-*end*	*What did she think he said __?*
…		

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Modeling acquisition

Because *wh*-dependencies are perceived as sequences of container nodes, local pieces of dependency structure can be characterized by container node trigrams.

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?
IP VP CP$_{null}$ IP VP PP

begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* =

begin-IP-VP
IP-VP-CP$_{null}$
VP-CP$_{null}$-IP
CP$_{null}$-IP-VP
IP-VP-PP
VP-PP-*end*

Pearl & Sprouse 2013a, 2013b, under review

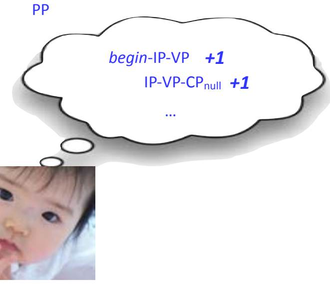
Subjacency-ish: Modeling acquisition

A child learns about the frequency of container node trigrams…

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?
IP VP CP$_{null}$ IP VP PP

begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* =

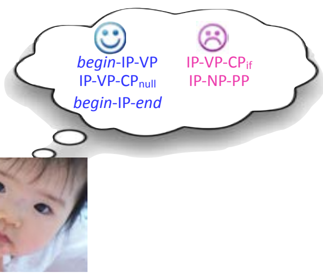
begin-IP-VP
IP-VP-CP$_{null}$
VP-CP$_{null}$-IP
CP$_{null}$-IP-VP
IP-VP-PP
VP-PP-*end*



Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Modeling acquisition

…and at the end of the learning period has a sense of the probability of any given container node trigram, based on its relative frequency.



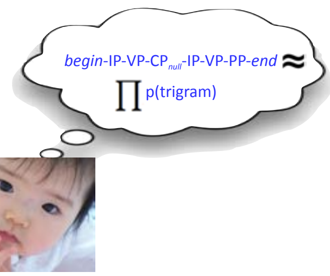
Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Modeling acquisition

A dependency can then have a probability, based on the product of the smoothed probabilities of its trigrams.

Who did she think the gift was from __?

Probability(*begin*-IP-VP-CP$_{null}$-IP-VP-PP-*end*)
= p(*begin*-IP-VP)
* p(IP-VP-CP$_{null}$)
* p(VP-CP$_{null}$-IP)
* p(CP$_{null}$-IP-VP)
* p(IP-VP-PP)
* p(VP-PP-*end*)



Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Modeling acquisition

This allows the modeled learner to generate judgments about the grammaticality of any dependency.

Higher probability dependencies are more grammatical while lower probability dependencies are less grammatical.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Realistic learning period

Children hear approximately one million utterances in their first three years (Hart & Risley 1995).

Assumption: learning period for modeled learners is 3 years (ex: between 2 and 5 years old for modeling children's acquisition of islands), so they would hear one million utterances.

Total learning period: 200,000 *wh*-dependency data points (*wh*-dependencies make up approximately 20% of the input)

Pearl & Sprouse 2013a, 2013b, under review

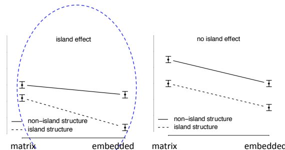
Subjacency-ish: Success metric

For each set of island stimuli from Sprouse et al. (2012), we generate grammaticality preferences for the modeled learner based on the dependency's probability.

	matrix	embedded
non-island	*begin-IP-end*	*begin-IP-VP-CP$_{that}$-IP-VP-end*
island	*begin-IP-end*	**begin-IP-VP-NP-CP$_{that}$-IP-VP-end*

We can then plot the log probability of the dependency on the y-axis of the interaction plot.

Non-parallel lines indicate superadditivity, which indicates knowledge of islands.

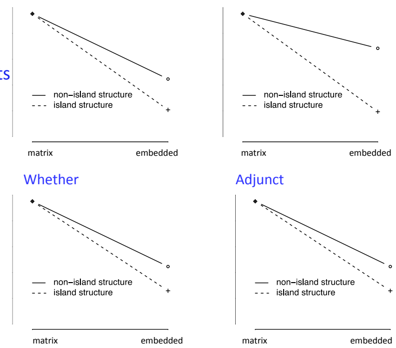


Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Success!

Superadditivity observed for all four islands — the qualitative behavior suggests that this learner has knowledge of these syntactic islands.

The subjacency-ish representation that relies on container node trigrams is useful for acquisition.



Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish: Take away

Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

Representation validation

If dependencies are represented as container node sequences, acquisition works well for these four syntactic islands. The learner can leverage probabilities of container node trigrams.

Pearl & Sprouse 2013a, 2013b, under review

Subjacency-ish vs. Subjacency: What's in UG?

Subjacency-ish

Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

Fewer pieces of knowledge necessarily in UG.

UG = innate + domain-specific

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & subcategorize by CP	?	?	*	
Low probability items are dispreferred	*			*

Subjacency

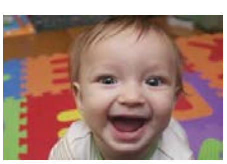
Wh … [$_{BN1}$ … [$_{BN2}$ … __]]

	Innate	Derived	Domain-specific	Domain-general
Attend to bounding nodes (BNs)	*		*	
Dependencies crossing 2+ BNs are not allowed	*		*	

Pearl & Sprouse 2013a, 2013b, under review

Syntactic islands: Big picture

This approach allows us to evaluate a representation of dependencies by using it for acquisition. We can then refine our theories of what must be in Universal Grammar.



Pearl & Sprouse 2013a, 2013b, under review

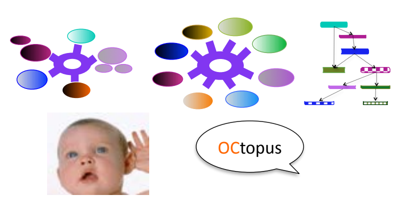
Computational acquisition modeling: Big picture

Informing theories of representation & acquisition

Metrical phonology:

❖ Can identify learning assumptions (like productive data filters) that benefit children using different knowledge representations

❖ Can identify language-specific grammars within these representations that are easier to learn from realistic data than the current versions



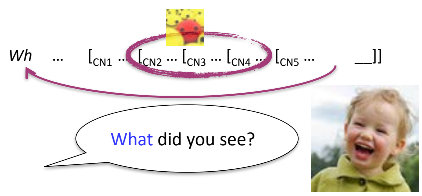
Computational acquisition modeling: Big picture

Informing theories of representation & acquisition

Syntax

❖ Can validate representations that make it easy to learn syntactic islands, and provide alternative proposals for what's in UG

❖ Can provide concrete demonstrations of learning strategies using these representations that succeed on realistic input data



Computational acquisition modeling: Big picture

This technique is a useful tool — so let's use it to inform our theories of representation and acquisition!

Pearl & Sprouse 2013b, Pearl 2014, Pearl & Sprouse under review

Thank you!

Computation of Language Laboratory
UC Irvine

Jon Sprouse Tim Ho Zephyr Detrano

Pranav Anand, Alex Clark, Joanna Lee, Joseph Nunn, Matt Wagers
Misha Becker, Sandy Chung, Jeff Lidz, Colin Phillips, Charles Yang
Bob Berwick, Bob Frank, Jim McCloskey, William Sakas
Adrian Brasoveanu, Norbert Hornstein, Armin Mester, Virginia Valian

GALANA selection committee

Audiences at:
- Berkeley Linguistics Society Annual Meeting 2014
- UC Santa Cruz Linguistics colloquium 2014
- Logic & Philosophy of Science 2013 colloquium, UC Irvine
- Institute for Mathematical Behavioral Sciences 2013, UC Irvine
- Johns Hopkins University Cognitive Science colloquium 2013
- New York University Linguistics colloquium 2012
- UMaryland Mayfest 2012
- Input & Syntactic Acquisition Workshop 2012

NSF

This work was supported in part by NSF grants BCS-0843896 and BCS-1347028.

Extra material

Three knowledge representations

Parametric systems

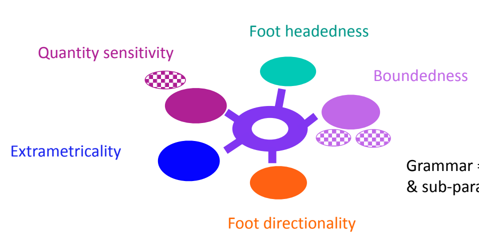
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Grammar = Set of parameter & sub-parameter values

Three knowledge representations

Parametric systems

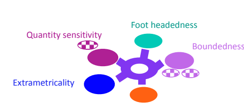
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



oc to pus

Three knowledge representations

Parametric systems

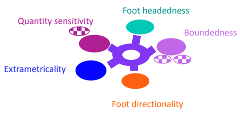
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



H L H
oc to pus

Quantity sensitivity
Are syllables all identical, or are they differentiated by syllable weight (into Heavy and Light syllables)?

Three knowledge representations

Parametric systems

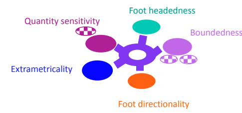
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



H L H
oc to pus

Extrametricality
Are all syllables included in the larger units of metrical feet, or are some excluded?

Three knowledge representations

Parametric systems

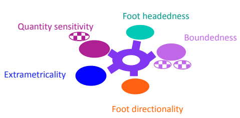
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



H L) H
oc to pus

Foot directionality
Are feet constructed from the left or from the right?

Three knowledge representations

Parametric systems

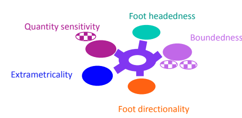
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



(H L) H
oc to pus

Boundedness
How big are metrical feet?

Three knowledge representations

Parametric systems

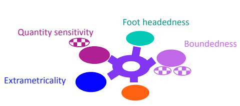
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



(H L) H
oc to pus

Foot headedness
Which syllable in a foot is stressed?

Three knowledge representations

Parametric systems

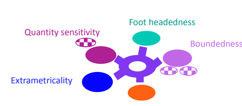
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



This grammar, comprised of particular parameter values, generates the correct stress contour.

(H L) H
OC to pus

Parameter values used:
Quantity sensitive, VC syllables = Heavy, Extrametricality on rightmost syllable, Feet built from the right, Foot = 2 syllables, Leftmost syllable in foot stressed

Three knowledge representations

Parametric systems

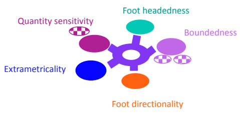
Correct grammar builds compatible contour

OCtopus

HV: Halle & Vergnaud 1987, Dresher 1999, Pearl 2011

5 parameters & 3 sub-parameters

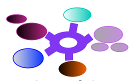
Hypothesis space: 156 grammars



This grammar, comprised of particular parameter values, generates the correct stress contour.

(H L) H
OC to pus

Parameter values used:
QS-VC-H, Em-Rt, FtDir-Rt, B-2-Syl, FtHd-Left



...which are the values of the English grammar.

Three knowledge representations

Parametric systems

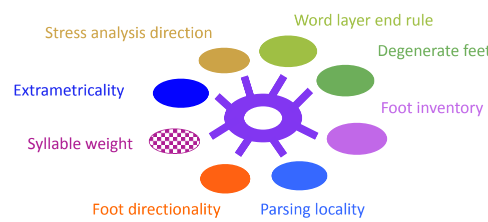
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Three knowledge representations

Parametric systems

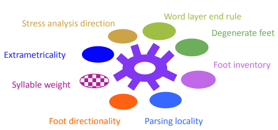
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



oc to pus

Three knowledge representations

Parametric systems

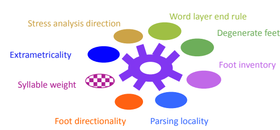
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



(...feet first...)

oc to pus

Stress analysis direction
Are metrical feet created before word-level stress is assigned to the edge syllables or after?

Three knowledge representations

Parametric systems

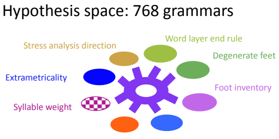
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



oc to pus

Extrametricality
Are syllables on the edge (or parts of syllables) excluded from metrical feet?

Three knowledge representations

Parametric systems

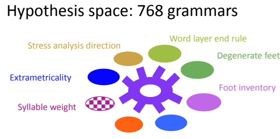
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



H L L
oc to pus

Syllable weight
Syllables are distinguished into Heavy and Light. Are syllables ending in VC (like *oc*) Heavy or Light?

Three knowledge representations

Parametric systems

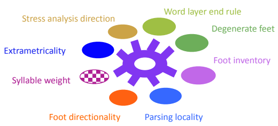
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



H L L)

oc to pus

Foot directionality

Are metrical feet constructed from the left or the right?

Three knowledge representations

Parametric systems

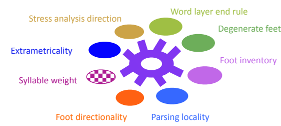
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



?

H L L

oc to pus

Parsing locality

Is one Light syllable skipped between metrical feet?

Three knowledge representations

Parametric systems

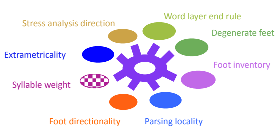
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



(H)(L L)

oc to pus

Foot inventory

How big are metrical feet?

Where does the stress fall within them?

Three knowledge representations

Parametric systems

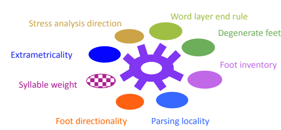
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



(H)(L L)

oc to pus

Degenerate feet

What do you do with leftover Light syllables if you have any?

Three knowledge representations

Parametric systems

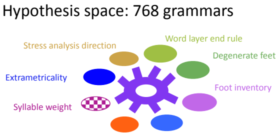
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



(H)(L L)

oc to pus

Word layer end rule

Where does word-level stress go if there are multiple stressed syllables? Can leftover Light syllables have word-level stress?

Three knowledge representations

Parametric systems

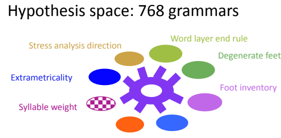
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



This grammar, comprised of particular parameter values, generates an incorrect stress contour.

(H)(L L)

OC TÓ pus

Parameter values used:

Bottom-up, Extrametricality on rightmost consonant, VC syllables = Heavy, Feet built from the right, Light syllables not skipped in between feet, Foot = Moraic trochee (2 moras with stress on leftmost), Single Light edge syllables not allowed to have stress, Rightmost syllable gets main stress

Three knowledge representations

Parametric systems

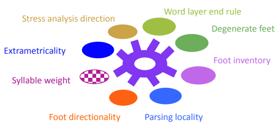
Correct grammar builds compatible contour

OCtopus

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars



This grammar, comprised of particular parameter values, generates an incorrect stress contour.

(H)(L L)

OC TÓ pus

Parameter values used:
Bot, Em-RtCons, VC-H, FtDir-Rt, PL-Strong, MorTro, DF-Strong, WLER-Rt

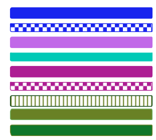
...which are the values of the English grammar.

Three knowledge representations

Constraint-ranking systems

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints



Three knowledge representations

Constraint-ranking systems

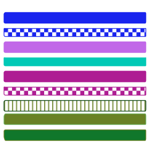
Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Premise: Many different candidates for a word's stress representation and contour are generated and then ranked according to which constraints are violated. Violating higher-ranked constraints is worse than violating lower-ranked constraints.



	C1	C2	C3	C4
(OC to) pus			*	*
oc (TO pus)	*		*	
(oc TO) pus		*	*	

Three knowledge representations

Constraint-ranking systems

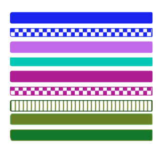
Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars



Grammar = ranked ordering of all constraints

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

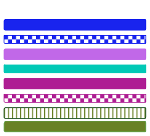
OCtopus

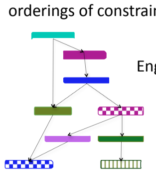
OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Official grammars for languages are often described as partial orderings of constraints.





English grammar

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

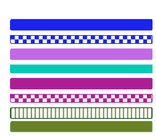
OCtopus

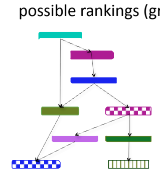
OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

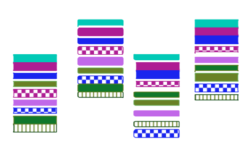
Hypothesis space: 9! rankings = 362,880 grammars

This means the "grammar" for a language is often a set of the possible rankings (grammars) that obey those orderings.





Ex: The English "grammar" is compatible with 26 rankings.



Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

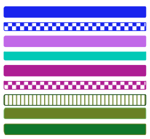
OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress



Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

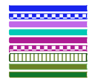
OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress



Nonfinality
Should the final syllable not be in a metrical foot?

(OC to) (PUS) — ✓ (OC to) pus

(oc TO) (PUS) — oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

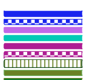
OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress



Parse-σ
Should all syllables be in metrical feet?

✓ (OC to) (PUS) — (OC to) pus

✓ (oc TO) (PUS) — oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

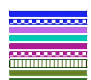
OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress



Foot binarity
Should all metrical feet consist of two units?

(OC to) (PUS) — ✓ (OC to) pus

(oc TO) (PUS) — ✓ oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Trochaic
Should metrical feet have stress on the leftmost syllable?

✓ (OC to) (PUS) — ✓ (OC to) pus

(oc TO) (PUS) — ✓ oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

BAby

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Weight-to-Stress (VV)
Should all VV syllables be stressed?

(ba BY) — ✓ (BA) (BY)

(BA) by — (BA by)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

Weight-to-Stress (VC)
Should all VC syllables be stressed?

✓ (OC to) (PUS) | (OC to) pus

(oc TO) (PUS) | oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

Align left
≈ Should metrical feet include the leftmost syllable?

✓ (OC to) (PUS) | ✓ (OC to) pus

✓ (oc TO) (PUS) | oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

Align right
≈ Should metrical feet include the rightmost syllable?

✓ (OC to) (PUS) | (OC to) pus

✓ (oc TO) (PUS) | ✓ oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

yourSELF

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

***Sonorant nucleus**
Should syllables not have sonorants (m, n, ŋ, l, r) as the nucleus?

✓ your (SELF) | (yr SELF)

✓ (YOUR) (SELF) | (YOUR slf)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

Sample candidates

A sample grammar that is a version of the English "grammar":

(OC to) (PUS) | (OC to) pus

(oc TO) (PUS) | oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Nonfinality, Parse-σ
Foot binarity
Trochaic
Weight-to-Stress
Align left, Align right
*Sonorant nucleus

Most important: Metrical feet have stress on the leftmost syllable.

Sample candidates

A sample grammar that is a version of the English "grammar":

(OC to) (PUS) | (OC to) pus

(oc TO) (PUS) | oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Next important: VV syllables are stressed.

A sample grammar that is a version of the English "grammar":

Sample candidates

(OC to) (PUS) (OC to) pus

(oc TO) (PUS) oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Next important: The final syllable is not included in a foot.

A sample grammar that is a version of the English "grammar":

Sample candidates

(OC to) (PUS) (OC to) pus

(oc TO) (PUS) oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Only one candidate left, and it has a compatible contour.

A sample grammar that is a version of the English "grammar":

Sample candidates

(OC to) (PUS) (OC to) pus

(oc TO) (PUS) oc (TO pus)

Three knowledge representations

Constraint-ranking systems

Best candidate for the correct grammar has a compatible contour

OCtopus

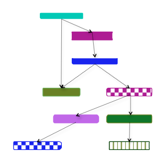
OT: Hammond 1999, Pater 2000, Tesar & Smolensky 2000

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

A sample grammar that is a version of the English "grammar":

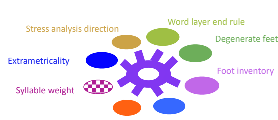


English "grammar"

Knowledge representation comparison

HV: 5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



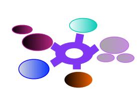
Hayes: 8 parameters

Hypothesis space: 768 grammars

OT: 9 violable constraints

Hypothesis space: 362,880 grammars

English instantiations

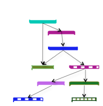


HV: 5 parameters & 4 sub-parameters

Hypothesis space: 156 grammars

Hayes: 8 parameters

Hypothesis space: 768 grammars

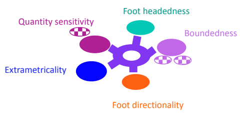


OT: 9 violable constraints

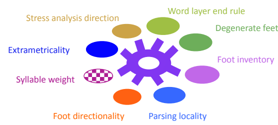
Hypothesis space: 362,880 grammars

(English = 26 grammars)

Knowledge representation comparison



HV: 5 parameters & 3 sub-parameters
Hypothesis space: 156 grammars

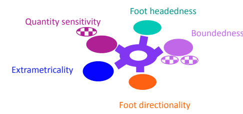


Hayes: 8 parameters
Hypothesis space: 768 grammars

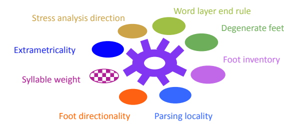
Each representation assumes certain syllabic distinctions.

OT: 9 violable constraints
Hypothesis space: 362,880 grammars

Knowledge representation comparison



HV: 5 parameters & 3 sub-parameters
Hypothesis space: 156 grammars
Syllabic distinctions: 3
(short, closed, long)



Hayes: 8 parameters
Hypothesis space: 768 grammars
Syllabic distinctions: 4
(short, potentially short, closed, long)

OT: 9 violable constraints
Hypothesis space: 362,880 grammars
Syllabic distinctions: 8
(short, sonorant, 4 closed variants, long, super-long)

Productive data filter

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).

Why would this occur?

Perhaps the learner realizes that some data are unproductive, and therefore likely irregular and unpredictable. The goal then becomes to learn a grammar that can account for all the data that are predictable.

Productive data filter

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).

How would this occur?

For every syllable word form (ex: V VV) that has multiple stress contours associated with it, the learner assumes that one of these patterns may be the productive contour and the others are exceptions to this general "rule".

Productive data filter

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).

How would this occur?

A formal way for identifying if there is a dominant rule for a set of items is the Tolerance Principle (Yang 2005, Legate & Yang 2012). This is used to estimate how many exceptions a rule can tolerate in a set before it's no longer useful for the learner to have the rule. If there are too many exceptions, it's better not to have a rule and learn patterns on an individual item basis instead of having a rule that keeps getting violated.

Productive data filter

Updated working assumption: The learner will try to learn a grammar that can account for all the productive data encountered (Legate & Yang 2012).

How would this occur?

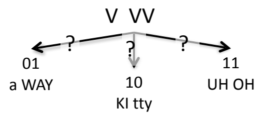
The number of exceptions a rule can tolerate for a set of N items is

$$\frac{N}{ln(N)}$$

(Yang 2005, Legate & Yang 2012)

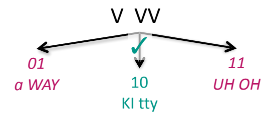
The Tolerance Principle in action

For every syllable word form with multiple stress contours, the learner could assess whether any of those contours is the dominant one (the "rule" for that syllable word form), using the Tolerance Principle.



The Tolerance Principle in action

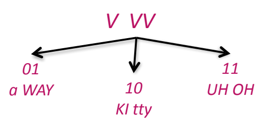
For every syllable word form with multiple stress contours, the learner could assess whether any of those contours is the dominant one (the "rule" for that syllable word form), using the Tolerance Principle.



If one contour is dominant, the learner should focus on accounting for that pattern, since it's regular and productive. The grammar should be able to generate it. The other contours can be ignored for purposes of learning the grammar.

The Tolerance Principle in action

For every syllable word form with multiple stress contours, the learner could assess whether any of those contours is the dominant one (the "rule" for that syllable word form), using the Tolerance Principle.



If no contour is dominant, the learner should ignore this syllable word form for the purposes of learning the grammar since there is no obvious regularity to account for.

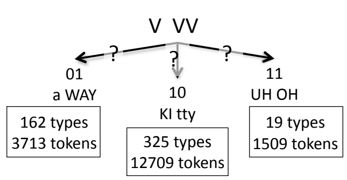
Productive data filter in action

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms
3 syllable distinctions

Hayes: 149 syllable word forms
4 syllable distinctions

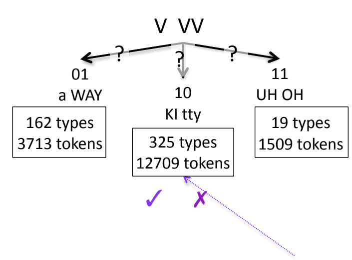


Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms
3 syllable distinctions

Hayes: 149 syllable word forms
4 syllable distinctions



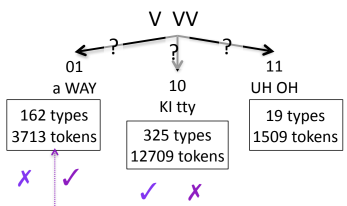
These items are good for the HV English grammar.

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



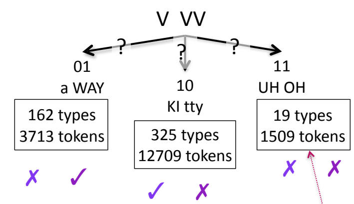
These items are good for the Hayes English grammar.

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



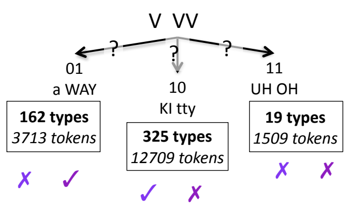
These items aren't good for either English grammar.

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



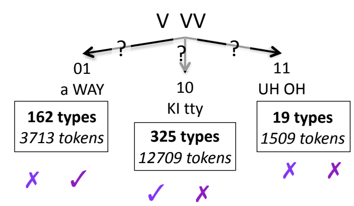
The Tolerance Principle looks at the **word types** with each stress pattern. Each represents an individual item that might follow the regular stress pattern rule (if there is one).

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



How many items should the stress "rule" apply to? N = 162 + 325 + 19 = 506

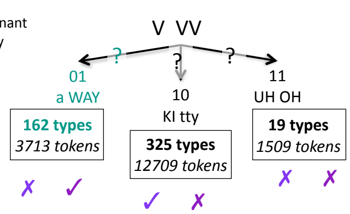
Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions

If this is the dominant pattern, too many exceptions:
325 + 19 > 81



How many exceptions are allowed? 506 / ln(506) = **81**

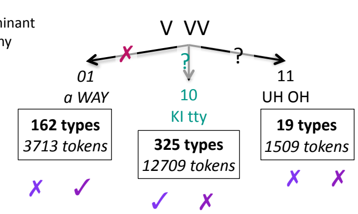
Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions

If this is the dominant pattern, too many exceptions:
162 + 19 > 81



How many exceptions are allowed? 506 / ln(506) = **81**

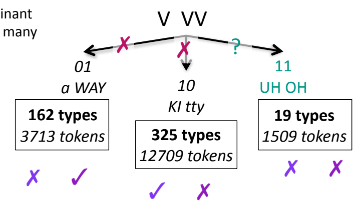
Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions

If this is the dominant pattern, way too many exceptions: 162 + 325 > 81



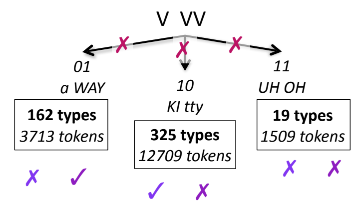
How many exceptions are allowed? 506 / ln(506) = **81**

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



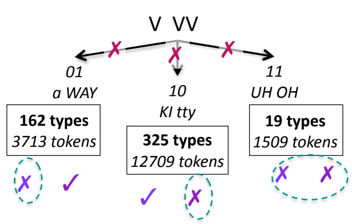
Learner conclusion: No dominant stress pattern, so none of these syllable word form data should be used to learn the English grammar.

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



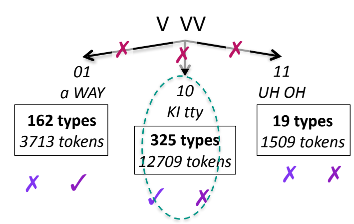
This will end up helping both grammars, since they won't be penalized for the patterns they can't account for.

Productive data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms — 3 syllable distinctions

Hayes: 149 syllable word forms — 4 syllable distinctions



However, the Hayes grammar is helped a little more, since it couldn't account for the most frequent stress pattern before, while the HV grammar could.

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

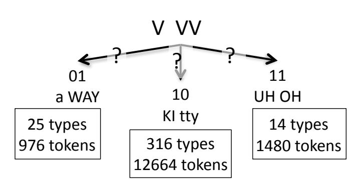
8 syllable distinctions

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions

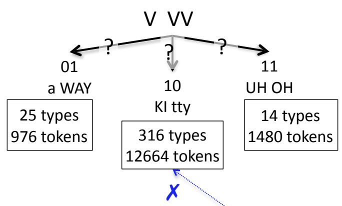


Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



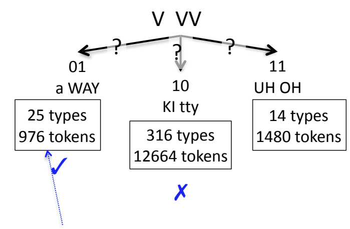
These items are bad for all English grammars.

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



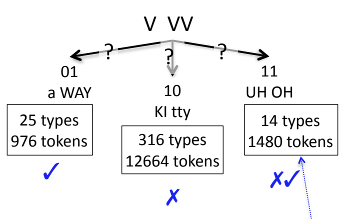
These items are good for most English grammars (21/26).

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



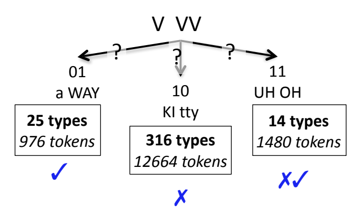
These items are good for a few English grammars (5/26).

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



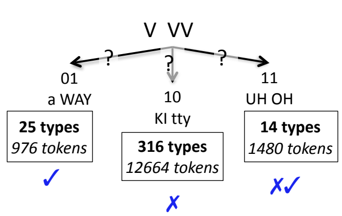
How many items should the stress "rule" apply to? N = 25 + 316 + 14 = 355

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



How many exceptions are allowed? 355 / ln(355) = **60**

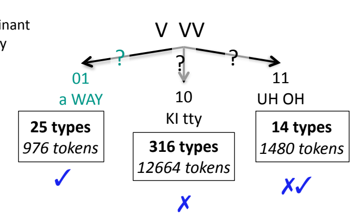
Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions

If this is the dominant pattern, too many exceptions: 316 + 14 > 60



How many exceptions are allowed? 355 / ln(355) = **60**

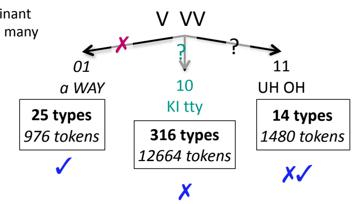
Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions

If this is the dominant pattern, NOT too many exceptions:
25 + 14 < 60



How many exceptions are allowed? 355 / ln(355) = **60**

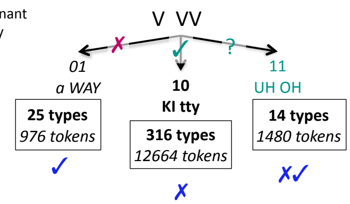
Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions

If this is the dominant pattern, too many exceptions:
25 + 316 > 60



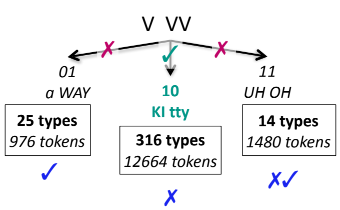
How many exceptions are allowed? 355 / ln(355) = **60**

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

452 syllable word forms

8 syllable distinctions



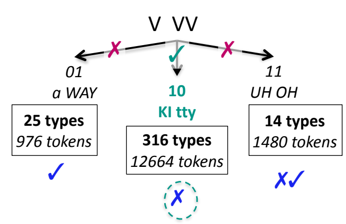
Under the OT syllable representation, there is a dominant stress pattern for this word form. Therefore, this pattern should be accounted for by the English grammar.

Productive data filter in action

Constraint-based: OT, with inflectional knowledge

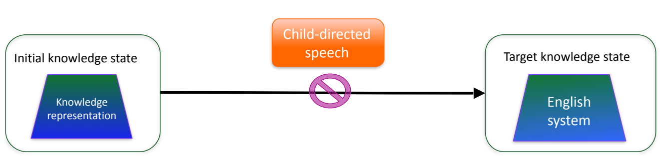
452 syllable word forms

8 syllable distinctions

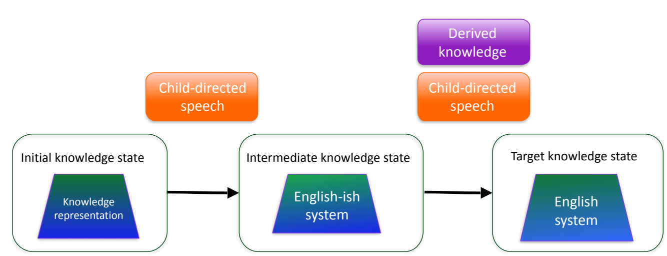


Unfortunately, this is the only pattern the English grammars cannot account for….this means a learner using the productivity filter would have even more trouble learning the current English OT grammar constraints.

The learnability problem



The learnability problem: One option



Change the (immediate) target state. Assume there is a transitory state in learning that the learner reaches and then leaves once additional knowledge is acquired.

Learning English metrical phonology

One solution: The learner has derived additional knowledge that helps guide learning.

General knowledge: Interactions with morphology
(Chomsky & Halle 1968, Hayes 1982, Kiparsky 1979)

Specific knowledge: Adding productive morphology doesn't change the stress pattern, even though all grammars base their stress patterns on the syllables present in the word.

EARly	PREtty	senSAtion
EARlier	PREttiest	senSAtional
		senSAtionally

Learning English metrical phonology

One solution: The learner has derived additional knowledge that helps guide learning.

English children seem to use inflectional morphology productively around 3 (Brown 1973) – so they may be aware it doesn't get stressed, based on their prior linguistic experience.

EARly	PREtty	senSAtion
EARlier	PREttiest	senSAtional
		senSAtionally

Learning English metrical phonology

So how does the (best) English grammar compare to the other grammars defined by the knowledge representation, once the learner knows inflectional morphology is stressless?

Relative compatibility of the English grammar = proportion of grammars in the hypothesis space the (best) English grammar is better than

Learning English metrical phonology

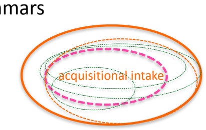
So how does the (best) English grammar compare to the other grammars defined by the knowledge representation, once the learner knows inflectional morphology is stressless?

Relative compatibility of the English grammar = proportion of grammars in the hypothesis space the (best) English grammar is better than

Relative compatibility of English grammar

HV: 0.712 by types out of 156 grammars
Hayes: 0.704 by types out of 768 grammars
OT: 0.786 by types out of 362,880 grammars

Better than many…but many are still better



Learning English metrical phonology

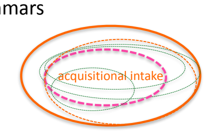
Implication:

There remain many other grammars in the hypothesis space that are more compatible with the data, even though the learner knows inflectional morphology is stressless. It would be easier to pick one of these other more compatible grammars.

Relative compatibility of English grammar

HV: 0.712 by types out of 156 grammars
Hayes: 0.704 by types out of 768 grammars
OT: 0.786 by types out of 362,880 grammars

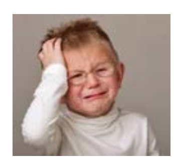
Better than many…but many are still better

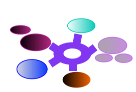


Learning English metrical phonology

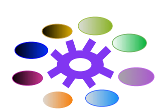
Continuing conclusion:

The same learnability issues persist for the English grammar in all three knowledge representations, even when the learner has some knowledge of the interactions between morphology and metrical phonology.

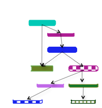




Parametric: HV



Parametric: Hayes



Constraint-based: OT

HV vs. Hayes on most frequent word forms

Stressed wordform	# Types	Examples	HV	Hayes
Lp	*592*	*water, doing, going*	*Yes*	*Yes*
Xp	*472*	*little, getting, coming*	*Yes*	*Yes*
Ll	*334*	*baby, sweetie, mommy*	*Yes*	*No*
Xl	*309*	*kitty, daddy, very*	*Yes*	*No*
Ap	*235*	*goodness, handsome, helper*	*Yes*	*Yes*
LL	*188*	*okay, byebye, TV*	*No*	*Yes*
Al	*172*	*window, birdie, only*	*Yes*	*No*
La	*171*	*peanuts, secrets, highest*	*Yes*	*No*
Xa	*170*	*biggest, buckets, hiccups*	*Yes*	*No*
xL	*145*	*below, today, hurray*	*No*	*Yes*

The impact of morphological knowledge

Example: What happens to words of the La stressed word form when the child gets morphological knowledge? (for the Hayes grammar, which can't account for it without morphological knowledge)

Stressed wordform	# Types	Examples	HV	Hayes
Lp	*592*	*water, doing, going*	*Yes*	*Yes*
Xp	*472*	*little, getting, coming*	*Yes*	*Yes*
Ll	*334*	*baby, sweetie, mommy*	*Yes*	*No*
Xl	*309*	*kitty, daddy, very*	*Yes*	*No*
Ap	*235*	*goodness, handsome, helper*	*Yes*	*Yes*
LL	*188*	*okay, byebye, TV*	*No*	*Yes*
Al	*172*	*window, birdie, only*	*Yes*	*No*
La	*171*	*peanuts, secrets, highest*	*Yes*	*No*
Xa	*170*	*biggest, buckets, hiccups*	*Yes*	*No*
xL	*145*	*below, today, hurray*	*No*	*Yes*

The impact of morphological knowledge

Example: What happens to words of the La stressed word form when the child gets morphological knowledge? (for the Hayes grammar, which can't account for it without morphological knowledge)

Before morphological knowledge
171 La (island, giant, moment)

After morphological knowledge
57 La (54 of the 171 + 3 added from Lp form)
– Hayes still can't account for these

100 Lp (father's→father pockets→pocket slobbered→slobber)
17 L (cutest→cute nicest→nice weirdest→weird)
- Hayes can now account for these

In this case, knowing inflectional morphology is stressless helps!

The impact of morphological knowledge

In general, the Hayes English grammars benefits from morphology knowledge (6.95% more types accounted for, due to 322 types), unlike the HV and OT English grammars.

Where are these changes happening?

- **28** types: incorrectly derived bisyllabics become monosyllabic
Examples: cleanest → clean (La→L); biggest → big (Xa→P); bestest → best (Aa→A)

- **100** types: incorrectly derived **La** becomes correctly-derived **Lp**
Examples: father's → father; pockets → pocket; slobbered → slobber

- **112** types: incorrectly derived **Xa** becomes correctly-derived **Xp**
Examples: sister's→ sister; apples→ apple; tickled→ tickle;

~ **92** types: Changes in less common wordforms
Examples: messages→ message; promises → promise; modeling → model

Proposed learning biases/capabilities

Several learning biases/capabilities are potentially both innate and domain-specific.

	Innate	Derived	Domain-specific	Domain-general
Learn from all *wh*-dependencies	?	?	*	
Parse data into phrase structure trees	?	?	*	
Attend to container nodes & subcategorize by CP	?	?	*	
Extract & track container node trigrams	*			*
Calculate dependency probability from trigrams	*			*

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Learn from all *wh*-dependencies	?	?	*	

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Learn from all *wh*-dependencies	?	?	*	

Clearly domain-specific, since this is language data.

May seem reasonable to attend to *wh*-dependency data when learning about *wh*-dependencies (and so this would be derived)

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Learn from all *wh*-dependencies	?	?	*	

Clearly domain-specific, since this is language data.

May seem reasonable to attend to *wh*-dependency data when learning about *wh*-dependencies (and so this would be derived)

…but then why not attend to *all* dependencies (ex: relative clause dependencies, binding dependencies) since *wh*-dependencies are a kind of dependency?

Empirical necessity of just using *wh*-dependency data:
There are different island effects for relative clauses (Sprouse et al. submitted) and no island effects for binding dependencies, so the learner needs to know to pay attention just to *wh*-dependencies.

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Parse data into phrase structure trees	?	?	*	

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Parse data into phrase structure trees	?	?	*	

Clearly domain-specific, since the structure is specific to language.

May be possible to bootstrap this information (acquiring syntactic categories: Mintz 2003, 2006; acquisition of hierarchical structure given syntactic categories as input: Klein & Manning 2002). If so, this would be derived…

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Parse data into phrase structure trees	?	?	*	

Clearly domain-specific, since the structure is specific to language.

May be possible to bootstrap this information (acquiring syntactic categories: Mintz 2003, 2006; acquisition of hierarchical structure given syntactic categories as input: Klein & Manning 2002). If so, this would be derived…

…but it's currently unclear if all the necessary phrase structure knowledge can be bootstrapped.

Important:
The need for this capability is not specific to learning islands – it's (presumably) needed for learning any kind of syntactic knowledge.

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & subcategorize by CP	?	?	*	

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & subcategorize by CP	?	?	*	

Identifying container nodes
- applies to language data: domain-specific
- derived from ability to parse utterances

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & subcategorize by CP	?	?	*	

Identifying container nodes
- applies to language data: domain-specific
- derived from ability to parse utterances

Attending to container nodes (among all the other data out there)
- applies to language data: domain-specific
- innate vs. derived?
  - could be specified innately (like bounding nodes)
  - could be derived from a bias to use representations that are already being used for parsing

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & **subcategorize by CP**	?	?	*	

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & **subcategorize by CP**	?	?	*	

About a linguistic representation: domain-specific

Innate vs. derived?
- Could be specified innately

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes & **subcategorize by CP**	?	?	*	

About a linguistic representation: domain-specific

Innate vs. derived?
- Could be specified innately
- Could be derived from prior linguistic experience:
  - Uncontroversial to assume children learn to distinguish different types of CPs since the lexical content of CPs has substantial consequences for the semantics of a sentence.
  - Also, adult speakers are sensitive to the distribution of *that* versus null complementizers (Jaeger 2010).

…but still have to know this is the right thing to subcategorize.

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Extract & track container node trigrams	*			*

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Extract & track container node trigrams	*			*

Applied in different cognitive domains: domain-general

Likely innate – learning with sequences of three units (transitional probabilities: Saffran et al. 1996, Aslin et al. 1998, Graf Estes et al. 2007, Pelucchi et al. 2009a, Pelucchi et al. 2009b; frequent frames for grammatical categorization: Mintz 2006, Wang & Mintz 2008)

...though why trigrams instead of some other n-gram?

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Calculate dependency probability from trigrams	*			*

Pearl & Sprouse 2013a, 2013b, under review

	Innate	Derived	Domain-specific	Domain-general
Calculate dependency probability from trigrams	*			*

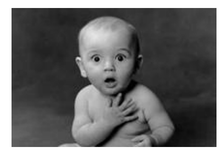
Applied in different cognitive domains: domain-general

Likely innate

Pearl & Sprouse 2013a, 2013b, under review

Main implications of this learner

(2) Even if Universal Grammar learning biases are required, they are different from (and less specific than) the biases previously proposed.



In particular, while one bias also specifies a particular linguistic representation, there is no bias defining the "constraint". This falls out from the other non-UG learning biases.

	Innate	Derived	Domain-specific	Domain-general
Learn from all *wh*-dependencies	?	?	*	
Attend to container nodes & subcategorize by CP	?	?	*	

vs.

	Innate	Derived	Domain-specific	Domain-general
Attend to bounding nodes (BNs)	*			*
Dependencies crossing 2+ BNs are not allowed	*			*

Pearl & Sprouse 2013a, 2013b, under review

Why learning from container node trigrams works

For each island-spanning dependency, there is at least one extremely low probability container node trigram in the dependency.

Complex NP island
start-IP-VP-NP-CP_{that}-IP-VP-*end*

Subject island
start-IP-VP-CP_{null}-IP-NP-PP-*end*

Whether island
start-IP-VP-$CP_{whether}$-IP-VP-*end*

Adjunct island
start-IP-VP-CP_{if}-IP-VP-*end*

These trigrams are never observed in the input – which is crucially different than being observed rarely. Thus, these islands are worse than dependencies involving trigrams that are rarely seen (e.g., dependencies with CP_{that}) and even longer dependencies that involve more frequent trigrams (e.g., triply embedded object dependencies using CP_{null}).

Pearl & Sprouse 2013a

The empirical necessity of trigrams

Not unigrams

A unigram model will successfully learn Whether and Adjunct islands, as there are container nodes in these dependencies that never appear in grammatical dependencies ($CP_{whether}$ and CP_{if})....but it will fail to learn Complex NP and Subject islands, as all of the container nodes in these islands are shared with grammatical dependencies.

Complex NP: *IP-VP-NP-CP_{that}-IP-VP
Subject: *IP-VP-CP_{null}-IP-NP-PP
Whether: IP-VP-$CP_{whether}$-IP-VP
Adjunct: IP-VP-CP_{if}-IP-VP

Pearl & Sprouse 2013a

The empirical necessity of trigrams

Not bigrams

At least for Subject islands, there is no bigram that occurs in a Subject island violation but not in any grammatical dependencies. The most likely candidate for such a bigram is IP-NP…However, sentences such as *What, again, about Jack impresses you?* or *What did you say about the movie scared you?* suggest that a gap can arise inside of NPs, as long as the extraction is of the head noun (what), not of the noun complement of the preposition.

Complex NP:	IP-VP-NP-CP_{that}-IP-VP
Subject:	*IP-VP-CP_{null}-IP-NP-PP
Whether:	IP-VP-$CP_{whether}$-IP-VP
Adjunct:	IP-VP-CP_{if}-IP-VP

Pearl & Sprouse 2013a

Parasitic gaps

The learner can't handle parasitic gaps, which are dependencies that span an island (and so should be ungrammatical) but which are somehow rescued by another dependency in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\_{true}$ [before reading $\_\_{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\_{parasitic}$] ultimately damage $\_\_{true}$?

Complex NP island

Pearl & Sprouse 2013a

Parasitic gaps

Why not? The current learner would judge the parasitic gap as ungrammatical since it is inside an island, irrespective of what other dependencies are in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\_{true}$ [before reading $\_\_{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\_{parasitic}$] ultimately damage $\_\_{true}$?

Complex NP island

This may be able to be addressed in a learner that is able to combine information from multiple dependencies in an utterance (perhaps because the learner has observed multiple dependencies resolved in utterances in the input).

Pearl & Sprouse 2013a

Across-the-board constructions

A similar problem occurs for across-the-board constructions.

Which book did you [[read __] and [then review __]]?
dependency for both gaps: IP-VP-VP

*Which book did you [[read the paper] and [then review __]]?
dependency for gap: IP-VP-VP

*Which book did you [[read __] and [then review the paper]]?
dependency for gap: IP-VP-VP

Again, this may be able to be addressed in a learner that is able to combine information from multiple dependencies in an utterance (perhaps because the learner has observed multiple dependencies resolved in utterances in the input).

Pearl & Sprouse 2013a

Some cross-linguistic issues

High probability trigrams that may be ungrammatical

Rizzi (1982): reports situations in Italian where simply doubling a grammatical sequence of trigrams leads to ungrammaticality…

IP-VP-CP_{wh}-IP-VP
but
*IP-VP-CP_{wh}-IP-VP-CP_{wh}-IP-VP-IP-VP

But these involve the same trigrams, so the learner in Pearl & Sprouse (forthcoming) will treat both the same (either grammatical or ungrammatical). If humans do have different judgments of these, then this cannot be accounted for by this learning algorithm.

Pearl & Sprouse 2013a

Complementizer *that*

That-trace effects

*Who do you think that __ read the book?
Who do you think __ read the book?

The current learning strategy captures this distinction.

Pearl & Sprouse 2013a

Complementizer *that*

That-trace effects

…but the current learning strategy will also generate a preference for object gaps without *that* compared to object gaps with *that.* (object *that*-trace effect)

What do you think that he read __ ? [prefers this one]
What do you think he read __ ?

Interestingly, Cowart 1997 finds an object *that-trace* effect, but it is much smaller than the subject *that-trace* effect

The model generates an asymmetrical dispreference when using adult-directed corpora, which contain more instances of *that* (5.40 versus 2.81). This could be taken to be a developmental prediction of the current algorithm: Children may disprefer object gaps in embedded *that-CP* clauses more than adults, and this dispreference will weaken as they are exposed to additional tokens of *that* in utterances containing dependencies.

Pearl & Sprouse 2013a

Now what?

This learning strategy for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow

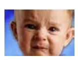
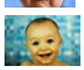
Pearl & Sprouse 2013a

Now what?

This learning strategy for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow

Subject extraction
*Who do you think ***that*** __ read the book?
Who do you think __ read the book?

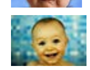
Pearl & Sprouse 2013a

Now what?

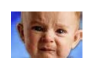
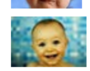
This learning strategy for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow

Subject extraction
*Who do you think ***that*** __ read the book?
Who do you think __ read the book?

Object extraction
What do you think ***that*** he read __ ?
What do you think he read __ ?

Pearl & Sprouse 2013a

Now what?

How does this learning strategy for *wh*-dependencies measure up cross-linguistically?

Island effects vary.
Ex: Italian does not have a subject island effect when the *wh*-dependency is part of a relative clause, though it does when the *wh*-dependency is part of a question. (Sprouse et al. submitted)

Would the input naturally lead our kind of learner to this distinction?

Now what?

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Related phenomena: The distribution of gaps

Parasitic gaps: Dependencies that span an island (and so should be ungrammatical) but which are somehow rescued by another dependency in the utterance.

*Which book did you laugh [before reading __]? Adjunct island
Which book did you judge __$_{true}$ [before reading __$_{parasitic}$]?

Pearl & Sprouse 2013a

Now what?

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Related phenomena: The distribution of gaps

Across-the-board (ATB) extraction: Similar situation.

Which book did you [[read __] and [then review __]]? — Coordinate structure island
dependency for both gaps: IP-VP-VP

*Which book did you [[read the paper] and [then review __]]?
dependency for gap: IP-VP-VP

*Which book did you [[read __] and [then review the paper]]?
dependency for gap: IP-VP-VP

Pearl & Sprouse 2013a

Now what?

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Semi-related phenomena: Binding dependencies

There don't appear to be the same restrictions on binding dependencies that there are on *wh*-dependencies.

The boy thought the joke about himself was really funny.

*Who did the boy think [the joke about __] was really funny? — Subject island

Pearl & Sprouse 2013a