More learnable than thou?
Empirically testing linguistic knowledge representations

Lisa Pearl
University of California, Irvine

Zephyr Detrano
Tim Ho

November 22, 2013:
Logic & Philosophy of Science Colloquium
University of California, Irvine
Knowledge

What exactly do you know when you know a language?

There are many types of linguistic knowledge, including knowledge of sounds (phonology)...
Knowledge

What exactly do you know when you know a language?

There are many types of linguistic knowledge, including knowledge of sounds (phonology), words (lexicon, morphology) ...
What exactly do you know when you know a language?

There are many types of linguistic knowledge, including knowledge of sounds (phonology), words (lexicon, morphology), sentences (syntax)…

Goblins steal children.

goblins

goblin + s = goblin (plural)

gáblínz

gabl
ínz
Knowledge

What exactly do you know when you know a language?

There are many types of linguistic knowledge, including knowledge of sounds (phonology), words (lexicon, morphology), sentences (syntax), meanings (semantics)...

goblins

$\text{goblin} + s = \text{goblin (plural)}$

gá blínz

g a b l í n z

steal(goblins, children)

Goblins steal children.
Knowledge

What exactly do you know when you know a language?

There are many types of linguistic knowledge, including knowledge of sounds (phonology), words (lexicon, morphology), sentences (syntax), meanings (semantics), and how language is used to convey certain types of information (pragmatics).

Don’t goblins steal children?

steal(goblins, children)

Goblins steal children.

goblins

goblin + s = goblin (plural)

gá blínz

g a b l í n z
Knowledge

What exactly do you know when you know a language?

Some things you can do when you know these different types of linguistic knowledge:

**Novel comprehension**: You can understand utterances in the language, even if you haven’t heard the specific utterance before.

**Novel generation**: You can create new utterances to express whatever thoughts you have.
Knowledge = Grammar

What exactly do you know when you know a language?

How can you do these things?

You have an underlying system that compactly describes the regularities and patterns of the language. This is your grammar for the language.
Grammar = Generative system

What exactly do you know when you know a language?

The grammar is often imagined as a generative system the speaker uses to comprehend and produce the language.
Generative system variation

Ideally, this generative system can be instantiated in various ways so that it can handle any of the world’s languages.
Grammar instantiations = Knowledge representation

What are the possible ways a grammar can be instantiated?

The knowledge representation encodes information about the general form that grammars for human languages can have.
Idea: If the child already knows the general form that grammars for human languages can have, all she needs to do is learn to instantiate her language’s grammar appropriately, based on the input data from her language.
Knowledge representations

What does a knowledge representation look like?

This demonstrates two checkpoints for any knowledge representation:

1. it should explain **cross-linguistic variation**
2. its language-specific grammar **should be learnable** from the data children encounter
Knowledge representations

What does a knowledge representation look like?

Argument from cross-linguistic variation:
Traditionally, proposals for the general form of knowledge representations have focused on optimizing the first checkpoint of accounting for cross-linguistic variation, with the (often implicit) assumption that the second checkpoint of learnability would be easily satisfied.
Knowledge representations

What does a knowledge representation look like?

Assumption of learnability:
If a child has the right knowledge representation, this makes learning the grammar of a language easy and fast.

Why? The child is just learning the specific instantiation, instead of having to figure out all the relevant variables from scratch.

Basic point: The right knowledge representation is helpful for acquisition (Chomsky 1981, Crain & Pietroski 2002, Dresher 1999)
Knowledge representations

What does a knowledge representation look like?

Argument from cross-linguistic variation:
This has led to several knowledge representations for the different aspects of linguistic knowledge. For example, proposals in metrical phonology have included both parametric and constraint-ranking systems.
Knowledge representations

What does a knowledge representation look like?

Argument from cross-linguistic variation:
These knowledge representations often overlap on some aspects of their representation components, but generally are not identical.

Parameters whose values must be set

Violable constraints that must be ranked
Knowledge representations

What does a knowledge representation look like?

When we have several proposals that satisfy the first criterion (accounting for cross-linguistic variation), it seems sensible to focus on the second criterion (learnability).
Knowledge representations

What does a knowledge representation look like?

Argument from acquisition: Is a language’s specific grammar in that knowledge representation learnable from the kind of data children of the language encounter?
Road map

Empirically grounding & quantifying learnability

Case study:
Knowledge representations in metrical phonology
- Knowledge representation comparison
- English as a tricky learning scenario
- Learnability results & implications
Road map

Empirically grounding & quantifying learnability

Case study:
Knowledge representations in metrical phonology
- Knowledge representation comparison
- English as a tricky learning scenario
- Learnability results & implications
Learnability

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
Learnability

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.
Learnability

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 perceived as relevant
To empirically ground a learnability analysis, we need to draw on a variety of methods.
To empirically ground a learnability analysis, we need to draw on a variety of methods.

Theoretical methods will define the knowledge representations, the set of grammars defined by a knowledge representation, and the language-specific grammar for a knowledge representation.
To empirically ground a learnability analysis, we need to draw on a variety of methods.

**Field work** methods will define the set of data that a language’s grammar should maximally account for, while **experimental** methods can define the data children are learning from.
To empirically ground a learnability analysis, we need to draw on a variety of methods.

Computational methods can analyze how much data any grammar defined by a knowledge representation can account for, including the one that’s intended to be that language’s grammar.
Quantifying learnability

Once we define the data set the child is learning from, 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.
Quantifying learnability

Once we define the data set the child is learning from, 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.

A grammar that can account for 70% of the data is better than a grammar that can only account for 55% of the data.
Quantifying learnability

Once we define the data set the child is learning from, 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.

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

Once we define the data set the child is learning from, 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. This indicates how easy it would be for a rational learner looking for the best grammar to choose it.

Example: A grammar with 1.00 relative compatibility is better than all other grammars defined by the knowledge representation.
Quantifying learnability

Once we define the data set the child is learning from, 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 (relative compatibility = 1.00) is compatible with. This is how much of the data that knowledge representation is capable of accounting for.

Example: If the best grammar can account for 70% of the data, this knowledge representation can account for 70% of the data at best.
Quantifying learnability

Once we define the data set the child is learning from, 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.

Working assumption:
The language-specific grammar should be the best grammar (relative compatibility = 1.00) for the data of that language, assuming a rational learner that’s looking for the best grammar.

It would be good if this grammar also had a high raw compatibility so that it would be useful to have, once learned.
Empirically grounding & quantifying learnability

Case study:

Knowledge representations in metrical phonology

- Knowledge representation comparison
- English as a tricky learning scenario
- Learnability results & implications
Case study:
A generative system of metrical phonology

Observable data: stress contour  OCtopus

Underlying representation determined by grammar?
Case study: A generative system of metrical phonology

Observable data: stress contour

Underlying representation determined by grammar?

Involves metrical feet:
Units larger than syllables but (often) smaller than words
Three knowledge representations

Parametric systems

Correct grammar builds compatible contour

OCtopus
Three knowledge representations

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Grammar = Set of parameter & sub-parameter values

Correct grammar builds compatible contour

OCtopus

- Quantity sensitivity
- Foot headedness
- Extrametricality
- Boundedness
- Foot directionality
Three knowledge representations

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Correct grammar builds compatible contour

*OCtopus*

- Quantity sensitivity
- Foot headedness
- Boundedness
- Extrametricality
- Foot directionality

`oc to pus`
Three knowledge representations

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Correct grammar builds compatible contour

Correct grammar builds compatible contour

Are syllables all identical, or are they differentiated by syllable weight (into Heavy and Light syllables)?
Three knowledge representations

Parametric systems

HV: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Correct grammar builds compatible contour

OCtopus

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

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999

5 parameters & 4 sub-parameters

Hypothesis space: 156 grammars

Correct grammar builds compatible contour

OCtopus

Are feet constructed from the left or from the right?
Three knowledge representations

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Correct grammar **builds** compatible contour

OCtopus

How big are metrical feet?
Three knowledge representations

Parametric systems

**HV:** Halle & Vergnaud 1987, Dresher 1999

5 parameters & 4 sub-parameters

Hypothesis space: 156 grammars

Correct grammar builds compatible contour

**OCtopus**

Which syllable in a foot is stressed?
Three knowledge representations

Parametric systems

**HV**: Halle & Vergnaud 1987, Dresher 1999
5 parameters & 4 sub-parameters
Hypothesis space: 156 grammars

Correct grammar builds compatible contour

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

\((H \quad L) \quad H\)

**OC**topus
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

OCtopus

Correct grammar builds compatible contour

Stress analysis direction
Extrametricality
Syllable weight
Foot directionality
Word layer end rule
Degenerate feet
Foot inventory
Parsing locality
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

Grammar = set of parameter values used to build a stress contour
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

OCtopus

Syllable weight
Extrametricality
Foot directionality
Word layer end rule
Degenerate feet
Foot inventory
Parsing locality
Stress analysis direction

oc to pus
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

OCtopus

(...feet first...)

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

oc to pus
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

Are syllables on the edge (or parts of syllables) excluded from metrical feet?
Three knowledge representations

Parametric systems

Hayes: Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Correct grammar builds compatible contour

OCtopus

Syllables are distinguished into Heavy and Light. Are syllables ending in VC (like oc) Heavy or Light?
Three knowledge representations

Parametric systems

**Hayes**: Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Are metrical feet constructed from the left or the right?
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

Are Light syllables skipped when building feet?
Three knowledge representations

Parametric systems

Hayes: Hayes 1995
8 parameters
Hypothesis space: 768 grammars

Correct grammar builds compatible contour

OCtopus

How big are metrical feet?
Where does the stress fall within them?
Three knowledge representations

Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Correct grammar builds compatible contour

OCtopus

What do you do with leftover Light syllables if you have any?
Three knowledge representations

Parametric systems

**Hayes**: Hayes 1995

8 parameters

Hypothesis space: 768 grammars

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

Correct grammar builds compatible contour

OCtopus

\((H)(L)\)

oc to pus
Three knowledge representations

Parametric systems

**Hayes**: Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Correct grammar *builds* compatible contour

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

\[(H)(L \, L)\]

**OCTopus**
Three knowledge representations

Constraint-ranking systems


9 violable constraints
Three knowledge representations

Constraint-ranking systems

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.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(OC to) pus</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>oc (TO pus)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(oc TO) pus</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Three knowledge representations

Constraint-ranking systems

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

9 violable constraints

Hypothesis space: $9!$ rankings = 362,880 grammars

Grammar = ranked ordering of all constraints

Best candidate for the correct grammar has a compatible contour
Three knowledge representations

Constraint-ranking systems

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

9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

**Best candidate** for the correct grammar has a compatible contour

**OCtopus**

Official grammars for actual languages are often described as partial orderings of constraints.
Three knowledge representations

Constraint-ranking systems

**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

**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

Best candidate for the correct grammar has a compatible contour

**OCtopus**

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

Constraint-ranking systems

**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*

---

**Best candidate** for the correct grammar has a compatible contour

**OCtopus**

Should the final syllable not be in a metrical foot?

- $(OC \text{ to}) (PUS)$
- $(oc \text{ TO}) (PUS)$
- $(OC \text{ to}) \text{ pus}$
- $oc \text{ (TO pus)}$
Three knowledge representations

Constraint-ranking systems

**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

---

Best candidate for the correct grammar has a compatible contour

**OCtopus**

Should all syllables be in metrical feet?

- $(OC \text{ to}) (PUS)$
- $(oc \text{ TO}) (PUS)$
- $(OC \text{ to}) \text{ pus}$
- $oc \text{ (TO pus)}$
Three knowledge representations

Constraint-ranking systems


9 violable constraints

Hypothesis space: 9! rankings = 362,880 grammars

Principle (Rooting): All words must have stress

Best candidate for the correct grammar has a compatible contour

OCtopus

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

**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

Best candidate for the correct grammar has a compatible contour

**OCtopus**

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

9 violable constraints
Hypothesis space: $9!$ rankings = 362,880 grammars
Principle (Rooting): All words must have stress

Best candidate for the correct grammar has a compatible contour

Should all VV syllables be stressed?

(\text{ba BY})
(\text{BA} \text{ by})
(\text{BA by})
Three knowledge representations

Constraint-ranking systems

**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

---

Best candidate for the correct grammar has a compatible contour

OCtopus

---

Should all VC syllables be stressed?

- (OC to) (PUS)
- (OC to) pus
- (oc TO) (PUS)
- oc (TO pus)
Three knowledge representations

**Constraint-ranking systems**

**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

---

**OCtopus**

Best candidate for the correct grammar has a compatible contour

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**

**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

Best candidate for the correct grammar has a compatible contour

OCR to PUS

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**

**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

Best candidate for the correct grammar has a compatible contour

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

✓

your (SELF)

✓

(your SELF)

✓

(YOUR) (SELF)

✓

(YOUR slf)
Three knowledge representations

Constraint-ranking systems

**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: (OC to) (PUS) (OC to) pus

Sample candidates:

- (OC to) (PUS)
- (OC to) pus
- (oc TO) (PUS)
- oc (TO pus)

Best candidate for the correct grammar has a compatible contour: OCtopus
Three knowledge representations

**Constraint-ranking systems**

**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

Best candidate for the correct grammar has a compatible contour

**OCtopus**

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

Sample candidates

A sample grammar:

- (OC to) (PUS)
- (OC to) pus
- (oc TO) (PUS)
- oc (TO pus)
Three knowledge representations

Constraint-ranking systems

**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.

Sample candidates

**A sample grammar:**

Best candidate for the correct grammar has a compatible contour

**OCtopus**
Three knowledge representations

Constraint-ranking systems

**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.

Sample candidates

A sample grammar:

Best candidate for the correct grammar has a compatible contour

**OCtopus**

Sample candidates:

- (OC to) (PUS)
- (OC to) pus
- (oc TO) (PUS)
- oc (TO pus)
Three knowledge representations

**Constraint-ranking systems**

**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: 

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

**Best candidate** for the correct grammar has a compatible contour

OCtopus

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

Sample candidates

- (OC to) (PUS)
- (OC to) pus
- (oc TO) (PUS)
- oc (TO pus)
Knowledge representation comparison

**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
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

This is generally a problem for acquisition.
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.
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
- EARlier
- PREtty
- PREttiest
- senSAtion
- senSAtional
- senSAtionally
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:


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

<table>
<thead>
<tr>
<th>NOUNS</th>
<th>VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONduct</td>
<td>conDUCT</td>
</tr>
<tr>
<td>DESert</td>
<td>deSERT</td>
</tr>
<tr>
<td>SUSpect</td>
<td>suSPECT</td>
</tr>
</tbody>
</table>
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: \( V \ VV \)

<table>
<thead>
<tr>
<th>KI tty</th>
<th>a WAY</th>
<th>UH OH</th>
</tr>
</thead>
<tbody>
<tr>
<td>V vv</td>
<td>v VV</td>
<td>V VV</td>
</tr>
</tbody>
</table>
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
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.

<table>
<thead>
<tr>
<th>Syllabic word form: $V$ $VV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select one of these...</td>
</tr>
<tr>
<td>KI tty</td>
</tr>
<tr>
<td>a WAY</td>
</tr>
<tr>
<td>UH OH</td>
</tr>
<tr>
<td>V vv</td>
</tr>
<tr>
<td>v VV</td>
</tr>
<tr>
<td>V VV</td>
</tr>
</tbody>
</table>

- Nonfinality, Parse-$\sigma$
- Foot binarity
- Trochaic
- Weight-to-Stress
- Align left, Align right
- *Sonorant nucleus
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: \( V \ VV \)

\[
\begin{align*}
\text{Kitty} & \quad \text{a WAY} & \quad \text{UH OH} \\
\text{vv} & \quad \text{v} & \quad \text{VV} \\
\text{VV} & \quad \text{V} & \quad \text{VV}
\end{align*}
\]
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: 95 of 186 syllabic word forms
Hayes: 86 of 149 syllabic word forms
OT: 166 of 452 syllabic word forms

This occurs a lot!
Learning English metrical phonology

So what’s the best any grammar in a given knowledge representation actually does, given these data?

Learnability potential = proportion of data the best grammar (relative compatibility = 1.00) can account for
Learning English metrical phonology

So what’s the best any grammar in a given knowledge representation actually does, given these 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 (0.739 tokens)
- Hayes: 0.683 types (0.750 tokens)
- OT: 0.657 types (0.729 tokens)

Around 2/3 of the word types
Learning English metrical phonology

Implication:
The best grammar 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 (0.739 tokens)
- **Hayes:** 0.683 types (0.750 tokens)
- **OT:** 0.657 types (0.729 tokens)

Around 2/3 of the word types

Data
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for

**Raw compatibility of English grammar**

<table>
<thead>
<tr>
<th>Method</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>0.593</td>
<td>0.716</td>
</tr>
<tr>
<td>Hayes</td>
<td>0.485</td>
<td>0.531</td>
</tr>
<tr>
<td>OT</td>
<td>0.573</td>
<td>0.574</td>
</tr>
</tbody>
</table>

Significantly less than the best grammar
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for

**Raw compatibility of English grammar**
- HV: 0.593 types (0.716 tokens)
- Hayes: 0.485 types (0.531 tokens)
- OT: 0.573 types (0.574 tokens)

Significantly less than the best grammar

**Raw compatibility of best grammar**
- HV: 0.668 types (0.739 tokens)
- Hayes: 0.683 types (0.750 tokens)
- OT: 0.657 types (0.729 tokens)
Learning English metrical phonology

Implication:
A rational learner would not pick the English grammar for any of these knowledge representations. It would pick the best grammar instead.

**Raw compatibility of English grammar**
- HV: 0.593 types (0.716 tokens)
- Hayes: 0.485 types (0.531 tokens)
- OT: 0.573 types (0.574 tokens)

**Significantly less than the best grammar**

**Raw compatibility of best grammar**
- HV: 0.668 types (0.739 tokens)
- Hayes: 0.683 types (0.750 tokens)
- OT: 0.657 types (0.729 tokens)
Learning English metrical phonology

So how does the (best) English grammar compare to the other grammars defined by the knowledge representation?

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?

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 by types (0.673 by tokens) out of 156 grammars
Hayes: 0.676 by types (0.685 by tokens) out of 768 grammars
OT: 0.817 by types (0.785 by tokens) out of 362,880 grammars

Better than many...but many are still better
Learning English metrical phonology

Implication:
There are many other grammars in the hypothesis space that are more compatible with the data. Even if children aren’t optimal learners, it would be easier to pick one of these other more compatible grammars.

Relative compatibility of English grammar
HV: 0.673 by types (0.673 by tokens) out of 156 grammars
Hayes: 0.676 by types (0.685 by tokens) out of 768 grammars
OT: 0.817 by types (0.785 by tokens) out of 362,880 grammars

Better than many…but many are still better
Interim conclusion:

**Learnability issues exist** for the English grammar in all three knowledge representations.
The learnability problem

Initial knowledge state
Generative system

Child-directed speech

Target knowledge state
English system
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**
- **EARlier**
- **PREtty**
- **PREttiest**
- **senSAtion**
- **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.
Learning English metrical phonology

So what’s the best any grammar in a given knowledge representation actually does, given these data and the knowledge that inflectional morphology is stressless?

Learnability potential = proportion of data the best grammar (relative compatibility = 1.00) can account for
Learning English metrical phonology

So what’s the best any grammar in a given knowledge representation actually does, given these data and the knowledge that inflectional morphology is stressless?

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

Raw compatibility of best grammar

HV: 0.662 types (0.738 tokens)
Hayes: 0.683 types (0.750 tokens)
OT: 0.677 types (0.749 tokens)

Still around 2/3 of the word types...
Learning English metrical phonology

Implication:
The best grammar is still 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. However, knowing inflectional morphology is stressless doesn’t seem to help it account for any more than it could before...

Raw compatibility of best grammar
HV: 0.662 types (0.738 tokens)
Hayes: 0.683 types (0.750 tokens)
OT: 0.677 types (0.749 tokens)

Still around 2/3 of the word types...
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data and the knowledge that inflectional morphology is stressless?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data and the knowledge that inflectional morphology is stressless?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for

**Raw compatibility of English grammar**

**HV:** 0.605 types (0.719 tokens)

**Hayes:** 0.550 types (0.552 tokens)

**OT:** 0.578 types (0.575 tokens)

Still significantly less than the best grammar
Learning English metrical phonology

So how does the (best) English grammar in a given knowledge representation do, given these data and the knowledge that inflectional morphology is stressless?

Raw compatibility of the English grammar = proportion of data the (best) English grammar can account for

Raw compatibility of English grammar
HV: 0.605 types (0.719 tokens)
Hayes: 0.550 types (0.552 tokens)
OT: 0.578 types (0.575 tokens)

Raw compatibility of best grammar
HV: 0.662 types (0.738 tokens)
Hayes: 0.683 types (0.750 tokens)
OT: 0.677 types (0.749 tokens)

Still significantly less than the best grammar
Learning English metrical phonology

Implication:
A rational learner would still not pick the English grammar for any of these knowledge representations, even with knowledge that inflectional morphology is stressless. It would pick the best grammar instead.

Raw compatibility of English grammar
HV: 0.605 types (0.719 tokens)
Hayes: 0.550 types (0.552 tokens)
OT: 0.578 types (0.575 tokens)

Raw compatibility of best grammar
HV: 0.662 types (0.738 tokens)
Hayes: 0.683 types (0.750 tokens)
OT: 0.677 types (0.749 tokens)
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 (0.673 by tokens) out of 156 grammars
Hayes: 0.704 by types (0.685 by tokens) out of 768 grammars
OT: 0.786 by types (0.777 by tokens) 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. Even if children aren’t optimal learners, it would be easier to pick one of these other more compatible grammars.

Relative compatibility of English grammar

HV: 0.712 by types (0.673 by tokens) out of 156 grammars
Hayes: 0.704 by types (0.685 by tokens) out of 768 grammars
OT: 0.786 by types (0.777 by tokens) out of 362,880 grammars

Better than many…but many are still better
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.
The learnability problem: One option that didn’t work

Additional knowledge that **inflectional morphology is stressless** in English didn’t seem to help. Why not?
The learnability problem: One option that didn’t work

One problem:

All English grammars generally want long syllables (VV syllable nucleus) to be stressed (though the HV parametric grammar allows some exceptions). However, many English words have long syllables that aren’t stressed. These remain problematic even with knowledge about inflectional morphology.

<table>
<thead>
<tr>
<th>proper names</th>
<th>diminutives</th>
</tr>
</thead>
<tbody>
<tr>
<td>EL mo</td>
<td>KI tty</td>
</tr>
<tr>
<td>MAN dy</td>
<td>DA ddy</td>
</tr>
<tr>
<td></td>
<td>BLAN kie</td>
</tr>
<tr>
<td></td>
<td>SWEE tie</td>
</tr>
<tr>
<td></td>
<td>DO ggie</td>
</tr>
<tr>
<td></td>
<td>SO ckie</td>
</tr>
</tbody>
</table>
The learnability problem: One option that didn’t work

One problem:

All English grammars generally want long syllables (VV syllable nucleus) to be stressed (though the HV parametric grammar allows some exceptions). However, many English words have long syllables that aren’t stressed. These remain problematic even with knowledge about inflectional morphology.

Proper names

EL mo
MAN dy

Diminutives

KI tty
DA ddy
BLAN kie

SWEE tie
DO ggie
SO ckie

Unstressed long syllables
The learnability problem: One option that didn’t work...

One solution: Additional knowledge

Perhaps children learn that the /i/ (“ee”) diminutive ending behaves like inflectional morphology. Then, syllables containing this ending are expected to be stressless.

![Diagram showing proper names and diminutives]

- Proper names: EL mo, MAN dy
- Diminutives: KI tty, SWEE tie, DA ddy, DO ggie, BLAN kie, SO ckie
- Unstressed long syllables
The learnability problem: One option that didn’t work

One solution: Additional knowledge

Perhaps children learn that proper names are a coherent (semantic) class that may have different stress properties. This is similar to recognizing that grammatical categories may have different effects on stress.

proper names

EL mo
MAN dy

diminutives

KI tty
DA ddy
BLAN kie

SWEE tie
DO ggie
SO ckie

unstressed long syllables
The learnability problem: Another option

Initial knowledge state: Generative system

Child-directed speech

Target knowledge state: English system
The initial state of the learner includes prior knowledge that helps the learner learn in a more sophisticated way.
What kind of prior knowledge?

Helpful prior knowledge: Learning biases

Pearl 2008: The HV parametric English grammar can be learned from child-directed speech (in principle) if children are biased to learn only from data perceived as unambiguous for a particular parameter value.

In addition, children must learn the parameter values in particular orders (obeying certain order constraints).
Unambiguous data filter

Previous working assumption: The learner will try to learn a generative system that can account for all the data encountered.
**Updated** working assumption: The learner will try to learn a generative system that can account for all the **unambiguous** data encountered.

This is a small subset of the available data which can be viewed as **maximally informative**.
Updated working assumption: The learner will try to learn a generative system that can account for all the unambiguous data encountered.

Why would this occur?

Perhaps the learner prioritizes data that are viewed as highly informative. The goal then becomes to learn a system that can account for all these data.
Updated working assumption: The learner will try to learn a generative system that can account for all the unambiguous data encountered.

How would this occur?

The learner may look for cues that signal a data point is unambiguous for a particular parameter value or constraint ordering (Pearl 2008, Pearl 2011). These cues may be derived from attempting to analyze a data point with the existing parametric/constraint-ordering options (Fodor 1998, Pearl 2007) – if only one parameter value or constraint ordering is present in successful analyses, this is a cue.
What kind of prior knowledge?

Helpful prior knowledge: **Learning biases**

Potentially helpful: A bias to learn only from data viewed as **regular** (Legate & Yang 2012).
Previous working assumption: The learner will try to learn a generative system that can account for all the data encountered.
Updated working assumption: The learner will try to learn a generative system that can account for all the regular data encountered.
Regular data filter

Updated working assumption: The learner will try to learn a generative system that can account for all the regular data encountered.

Why would this occur?

Perhaps the learner realizes that some data are irregular and therefore unpredictable. The goal then becomes to learn a system that can account for all the data that are regular and therefore predictable.
Updated working assumption: The learner will try to learn a generative system that can account for all the regular data encountered.

How would this occur?

For every syllable word form (ex: V VV) that has multiple stress patterns associated with it, the learner assumes that one of these patterns may be the regular pattern and the others are exceptions to this general “rule”.
Updated working assumption: The learner will try to learn a generative system that can account for all the regular data encountered.

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.
Updated working assumption: The learner will try to learn a generative system that can account for all the regular data encountered.

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 patterns, the learner could assess **whether any of those patterns is the dominant pattern** (the “rule” for that syllable word form), using the Tolerance Principle.

![Diagram showing stress patterns in words]
The Tolerance Principle in action

For every syllable word form with multiple stress patterns, the learner could assess *whether any of those patterns is the dominant pattern* (the “rule” for that syllable word form), using the Tolerance Principle.

If one pattern is dominant, the learner should focus on accounting for that pattern, since it’s regular. The underlying generative system should be able to generate it. The other patterns can be ignored for purposes of learning the system.
The Tolerance Principle in action

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

If no pattern is dominant, the learner should ignore this syllable word form for the purposes of learning the underlying generative system since there is no obvious regularity to account for.
The learnability problem: A third option

Initial knowledge state

Generative system

Child-directed speech

Target knowledge state

English system
The learnability problem: A third option

The target grammar is different than we think, and we should update our theories about the definition of the English grammar.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Parametric: HV

- Quantity sensitivity
- Extrametricality
- Foot headedness
- Boundedness
- Foot directionality
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

If we examine the grammars with high compatibility, it turns out that the values used by the English grammar are the values used by the majority of these grammars.

Example: Extrametricality on the rightmost syllable is used by 53 of 58 high compatibility grammars.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

If we examine the grammars with high compatibility, it turns out that the values used by the English grammar are the values used by the majority of these grammars.

Example: Extrametricality on the rightmost syllable is used by 53 of 58 high compatibility grammars.

Upshot: Unclear for the HV knowledge representation that the learning problem can be fixed by simply switching one parameter value here or there.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Parametric: Hayes

- Stress analysis direction
- Extrametricality
- Syllable weight
- Foot directionality
- Word layer end rule
- Degenerate feet
- Foot inventory
- Parsing locality
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.

(1) Change the Foot inventory value.

This boosts compatibility from 0.485 types (0.531 tokens) to 0.644 types (0.733 tokens).

Relative compatibility = 0.910
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Parametric: Hayes

If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.

1. Change the Foot inventory value.
2. Change the Extrametricality and Degenerate feet values.

This boosts compatibility from 0.485 types (0.531 tokens) to 0.652 types (0.729 tokens).

Relative compatibility = 0.923
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Parametric: Hayes

If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.

(1) Change the Foot inventory value.
(2) Change the Extrametricality and Degenerate feet values.

Upshot: For the Hayes knowledge representation, the learning problem could be ameliorated by simply switching a small number of parameter values.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

**Constraint-based: OT**

- Nonfinality, Parse-σ
- Foot binarity
- Trochaic
- Weight-to-Stress
- Align left, Align right
- *Sonorant nucleus
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

If we examine the grammars with high compatibility, it turns out that there is one constraint ordering English uses which none of these grammars use.

**Weight-to-Stress VV >> Nonfinality**

(Effect: Prefer long syllables to be stressed, even if they’re at the right edge of the word)

_Baby_ problematic
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Constraint-based: OT

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

If we examine the grammars with high compatibility, it turns out that there is one constraint ordering English uses which none of these grammars use.

Implication: This ranking may be better for English data.
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Constraint-based: OT

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

If we examine the grammars with high compatibility, it turns out that there is one constraint ordering English uses which none of these grammars use.

Best English grammar compatibility with the original ranking is 0.574 types (0.573 tokens).

Best English grammar compatibility with this swapped ranking is 0.655 types (0.729 tokens).

**Relative compatibility = 0.988**

Optimal: 0.657 types (0.729 tokens)
What should the target grammar be?

One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

Constraint-based: OT

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

If we examine the grammars with high compatibility, it turns out that there is one constraint ordering English uses which none of these grammars use.

Upshot: For the OT knowledge representation, the learning problem could be alleviated by simply switching one constraint ordering.
Learnability implication recap

Since learnability issues exist for all three knowledge representations, something else must be going on.

As it stands, they can explain cross-linguistic variation, but the English grammars don’t seem to be learnable from English data.
Learnability implication recap

One thing we can do to help learnability:

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

Adding some knowledge about the interaction of stress with inflectional morphology is insufficient on its own, however.
Learnability implication recap

Another thing we can do to help learnability:

Add in **prior knowledge** that helps the child perceive the input in a helpful way. Two potentially useful learning biases involve learning a system that accounts for a subset of the available data, rather than all of it. One parametric system (**HV**) has been shown to be learnable this way.
A third thing we can do to help learnability:

Update our ideas about the **target grammar** slightly (alter certain parameter values or constraint rankings).
Big picture
Establishing a methodology for quantitatively evaluating competing linguistic knowledge representations

• based on an argument from acquisition (how learnable are they from realistic data?)
Establishing a methodology for quantitatively evaluating competing linguistic knowledge representations

- based on an argument from acquisition (how learnable are they from realistic data?)

Testing against hard cases, like learning English

Idea: If a representation can handle the hard cases, it can handle anything
Establishing a methodology for quantitatively evaluating competing linguistic knowledge representations

- based on an argument from acquisition (how learnable are they from realistic data?)

Testing against hard cases, like learning English

Idea: If a representation can handle the hard cases, it can handle anything

Provide insight on why current instantiations fail and what can be done about it

- alter theory about how learning proceeds

- alter theory of language-specific grammars defined by knowledge representations
Thank you!

Zephyr Detrano  

Tim Ho

Computation of Language Laboratory
UC Irvine
Extra material
Knowledge representation comparison

Each representation assumes certain syllabic distinctions.

**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
Knowledge representation comparison

**HV**: 5 parameters & 4 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)
Learning English metrical phonology: Non-trivial

So what’s the best any grammar could possibly do, given these data?

Answer: Syllabic learnability potential

Account for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV

- KItty
- a WAY
- UH OH
- V vv
- v VV
- V VV

Data
Learning English metrical phonology: Non-trivial

So what’s the best any grammar could possibly do, given these data?

Answer: Syllabic learnability potential

Account for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV

Syllabic learnability potential by knowledge representation (proportion accounted for)

<table>
<thead>
<tr>
<th></th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>0.711</td>
<td>0.766</td>
</tr>
<tr>
<td>Hayes</td>
<td>0.719</td>
<td>0.769</td>
</tr>
<tr>
<td>OT</td>
<td>0.795</td>
<td>0.829</td>
</tr>
</tbody>
</table>

Pretty good potential coverage, even if it isn’t perfect
Learning English metrical phonology

So what’s the best any grammar could possibly do, given these data and the knowledge that inflectional morphology is stressless?

Syllabic learnability potential

Accounting for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV

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

data
Learning English metrical phonology

So what’s the best any grammar could possibly do, given these data and the knowledge that inflectional morphology is stressless?

Syllabic learnability potential
Accounting for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV

Syllabic learnability potential by knowledge representation (proportion accounted for)

<table>
<thead>
<tr>
<th></th>
<th>HV: 0.708 types (0.766 tokens)</th>
<th>Hayes: 0.711 types (0.768 tokens)</th>
<th>OT: 0.820 types (0.852 tokens)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Still pretty good potential coverage...but it’s also still not perfect.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Learning English metrical phonology

Why wouldn’t the best grammar’s compatibility be equivalent to the syllabic learnability potential? What’s preventing it?

A grammar is constrained by the specific parameters or constraints defined by the KR. These parameters/constraints may not be compatible with the most frequent stress contour every time.

Syllabic word form:  V  VV

| Kl tty | a  WAY |
| V vv | v  VV |
| 18 | 242 |

Syllabic word form:  VV  VV

| BA by | o  KAY |
| VV vv | vv  VV |
| 198 | 4 |

Syllabic learnability potential: Choose these stress contours (which are opposite)
Learning English metrical phonology

Why wouldn’t the best grammar’s compatibility be equivalent to the syllabic learnability potential? What’s preventing it?

A grammar is constrained by the specific parameters or constraints defined by the KR. These parameters/constraints may not be compatible with the most frequent stress contour every time.

Syllabic word form: V VV

Syllabic word form: VV VV

Grammar compatibility (preferring same contour): Choose these stress contours...
Learning English metrical phonology

Why wouldn’t the best grammar’s compatibility be equivalent to the syllabic learnability potential? What’s preventing it?

A grammar is constrained by the specific parameters or constraints defined by the KR. These parameters/constraints may not be compatible with the most frequent stress contour every time.

Syllabic word form:  V  VV

V vv
18

VV vv
198

Syllabic word form:  VV  VV

VV vv
242

V V

18

BA by
4

198

o  KAY

Grammar compatibility (preferring same contour): ...or these stress contours.
Parametric knowledge representation comparison
• HV English grammar outperforms Hayes English grammar in overall compatibility

Raw compatibilities by type for English grammar in parameter systems (with no knowledge of inflection):

<table>
<thead>
<tr>
<th></th>
<th>HV</th>
<th>Hayes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>59.33%</td>
<td>48.05%</td>
</tr>
</tbody>
</table>
Which common stressed wordforms are contributing to the relatively high performance of the HV English grammar?

The 10 most common stressed wordforms by type constitute 58.32% of total types in the input.

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp</td>
<td>592</td>
<td>water, doing, going</td>
</tr>
<tr>
<td>Xp</td>
<td>472</td>
<td>little, getting, coming</td>
</tr>
<tr>
<td>Ll</td>
<td>334</td>
<td>baby, sweetie, mommy</td>
</tr>
<tr>
<td>Xl</td>
<td>309</td>
<td>kitty, daddy, very</td>
</tr>
<tr>
<td>Ap</td>
<td>235</td>
<td>goodness, handsome, helper</td>
</tr>
<tr>
<td>LL</td>
<td>188</td>
<td>okay, byebye, TV</td>
</tr>
<tr>
<td>Al</td>
<td>172</td>
<td>window, birdie, only</td>
</tr>
<tr>
<td>La</td>
<td>171</td>
<td>peanuts, secrets, highest</td>
</tr>
<tr>
<td>Xa</td>
<td>170</td>
<td>biggest, buckets, hiccups</td>
</tr>
<tr>
<td>xL</td>
<td>145</td>
<td>below, today, hurray</td>
</tr>
</tbody>
</table>
HV English grammar derives 8 of the 10 most common stressed wordforms.

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
<th>HV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp</td>
<td>592</td>
<td>water, doing, going</td>
<td>Yes</td>
</tr>
<tr>
<td>Xp</td>
<td>472</td>
<td>little, getting, coming</td>
<td>Yes</td>
</tr>
<tr>
<td>Ll</td>
<td>334</td>
<td>baby, sweetie, mommy</td>
<td>Yes</td>
</tr>
<tr>
<td>Xl</td>
<td>309</td>
<td>kitty, daddy, very</td>
<td>Yes</td>
</tr>
<tr>
<td>Ap</td>
<td>235</td>
<td>goodness, handsome, helper</td>
<td>Yes</td>
</tr>
<tr>
<td>LL</td>
<td>188</td>
<td>okay, byebye, TV</td>
<td>No</td>
</tr>
<tr>
<td>Al</td>
<td>172</td>
<td>window, birdie, only</td>
<td>Yes</td>
</tr>
<tr>
<td>La</td>
<td>171</td>
<td>peanuts, secrets, highest</td>
<td>Yes</td>
</tr>
<tr>
<td>Xa</td>
<td>170</td>
<td>biggest, buckets, hiccups</td>
<td>Yes</td>
</tr>
<tr>
<td>xL</td>
<td>145</td>
<td>below, today, hurray</td>
<td>No</td>
</tr>
</tbody>
</table>
**Hayes English grammar** derives 5 of the 10 most common stressed wordforms.

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
<th>Hayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp</td>
<td>592</td>
<td><em>water, doing, going</em></td>
<td>Yes</td>
</tr>
<tr>
<td>Xp</td>
<td>472</td>
<td><em>little, getting, coming</em></td>
<td>Yes</td>
</tr>
<tr>
<td>Ll</td>
<td>334</td>
<td><em>baby, sweetie, mommy</em></td>
<td>No</td>
</tr>
<tr>
<td>Xi</td>
<td>309</td>
<td><em>kitty, daddy, very</em></td>
<td>No</td>
</tr>
<tr>
<td>Ap</td>
<td>235</td>
<td><em>goodness, handsome, helper</em></td>
<td>Yes</td>
</tr>
<tr>
<td>LL</td>
<td>188</td>
<td><em>okay, byebye, TV</em></td>
<td>Yes</td>
</tr>
<tr>
<td>Al</td>
<td>172</td>
<td><em>window, birdie, only</em></td>
<td>No</td>
</tr>
<tr>
<td>La</td>
<td>171</td>
<td><em>peanuts, secrets, highest</em></td>
<td>No</td>
</tr>
<tr>
<td>Xa</td>
<td>170</td>
<td><em>biggest, buckets, hiccups</em></td>
<td>No</td>
</tr>
<tr>
<td>xL</td>
<td>145</td>
<td><em>below, today, hurray</em></td>
<td>Yes</td>
</tr>
</tbody>
</table>
The **Hayes English grammar** misses **Ll** and **Xl**, two common patterns that together account for **13%** of the input by type.

However, it derives **LL** and **xL**, alternate contours for these wordforms that are also frequent, accounting for **7%** of input by type.

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
<th>HV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lp</strong></td>
<td>592</td>
<td><em>water, doing, going</em></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Xp</strong></td>
<td>472</td>
<td><em>little, getting, coming</em></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Ll</strong></td>
<td>334</td>
<td><em>baby, sweetie, mommy</em></td>
<td>No</td>
</tr>
<tr>
<td><strong>Xl</strong></td>
<td>309</td>
<td><em>kitty, daddy, very</em></td>
<td>No</td>
</tr>
<tr>
<td><strong>Ap</strong></td>
<td>235</td>
<td><em>goodness, handsome, helper</em></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>LL</strong></td>
<td>188</td>
<td><em>okay, byebye, TV</em></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Al</strong></td>
<td>172</td>
<td><em>window, birdie, only</em></td>
<td>No</td>
</tr>
<tr>
<td><strong>La</strong></td>
<td>171</td>
<td><em>peanuts, secrets, highest</em></td>
<td>No</td>
</tr>
<tr>
<td><strong>Xa</strong></td>
<td>170</td>
<td><em>biggest, buckets, hiccups</em></td>
<td>No</td>
</tr>
<tr>
<td><strong>xL</strong></td>
<td>145</td>
<td><em>below, today, hurray</em></td>
<td>Yes</td>
</tr>
</tbody>
</table>
# HV vs. Hayes on most frequent word forms

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
<th>HV</th>
<th>Hayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp</td>
<td>592</td>
<td>water, doing, going</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Xp</td>
<td>472</td>
<td>little, getting, coming</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Li</td>
<td>334</td>
<td>baby, sweetie, mommy</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Xi</td>
<td>309</td>
<td>kitty, daddy, very</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Ap</td>
<td>235</td>
<td>goodness, handsome, helper</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LL</td>
<td>188</td>
<td>okay, byebye, TV</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Al</td>
<td>172</td>
<td>window, birdie, only</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>La</td>
<td>171</td>
<td>peanuts, secrets, highest</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Xa</td>
<td>170</td>
<td>biggest, buckets, hiccups</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>xL</td>
<td>145</td>
<td>below, today, hurray</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
The impact of morphological knowledge

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

<table>
<thead>
<tr>
<th>Stressed wordform</th>
<th># Types</th>
<th>Examples</th>
<th>HV</th>
<th>Hayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lp</td>
<td>592</td>
<td>water, doing, going</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Xp</td>
<td>472</td>
<td>little, getting, coming</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ll</td>
<td>334</td>
<td>baby, sweetie, mommy</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Xl</td>
<td>309</td>
<td>kitty, daddy, very</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Ap</td>
<td>235</td>
<td>goodness, handsome, helper</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LL</td>
<td>188</td>
<td>okay, byebye, TV</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Al</td>
<td>172</td>
<td>window, birdie, only</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>La</td>
<td>171</td>
<td>peanuts, secrets, highest</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Xa</td>
<td>170</td>
<td>biggest, buckets, hicups</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>xL</td>
<td>145</td>
<td>below, today, hurray</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
The impact of morphological knowledge

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

Pre Morphology
171 La (island, giant, moment)

Post Morphology
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, inflectional morphology helps!
The impact of morphological knowledge

In general, the Hayes English 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
Regular data filter in action
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
    3 syllable distinctions

Hayes: 185 syllable words forms
    4 syllable distinctions

V   VV

01  a WAY
   162 types
   3713 tokens

10  Ktty
   325 types
   12709 tokens

11  UH OH
   19 types
   1509 tokens
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge
HV: 148 syllable word forms  
Hayes: 185 syllable words forms
  3 syllable distinctions  
  4 syllable distinctions

These items are good for the HV English grammar.
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
   3 syllable distinctions

Hayes: 185 syllable words forms
   4 syllable distinctions

These items are good for the Hayes English grammar.
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
3 syllable distinctions

Hayes: 185 syllable words forms
4 syllable distinctions

These items aren’t good for either English grammar.
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).
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
Hayes: 185 syllable words forms
3 syllable distinctions
4 syllable distinctions

V VV

01 a WAY
162 types
3713 tokens
X ✓

10 Ktty
325 types
12709 tokens
✓ X

11 UH OH
19 types
1509 tokens
X X

How many items should the stress “rule” apply to? N = 162 + 325 + 19 = 506
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
3 syllable distinctions

Hayes: 185 syllable words forms
4 syllable distinctions

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

How many exceptions are allowed? \( \frac{506}{\ln(506)} = 81 \)
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge
HV: 148 syllable word forms
Hayes: 185 syllable words forms
3 syllable distinctions
4 syllable distinctions

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

How many exceptions are allowed? $\frac{506}{\ln(506)} = 81$
Regular data filter in action

**Parametric: HV & Hayes, with inflectional knowledge**

HV: 148 syllable word forms  
Hayes: 185 syllable words forms  
3 syllable distinctions  
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
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge

HV: 148 syllable word forms
3 syllable distinctions

Hayes: 185 syllable words forms
4 syllable distinctions

Learner conclusion: No dominant stress pattern, so none of these syllable word form data should be used to learn the underlying generative system.
Regular data filter in action

**Parametric: HV & Hayes, with inflectional knowledge**

HV: 148 syllable word forms
3 syllable distinctions

Hayes: 185 syllable words 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.
Regular data filter in action

Parametric: HV & Hayes, with inflectional knowledge
HV: 148 syllable word forms
   3 syllable distinctions
Hayes: 185 syllable words 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.
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

406 syllable word forms
8 syllable distinctions
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

406 syllable word forms
8 syllable distinctions
Regular data filter in action

Constraint-based: **OT**, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

These items are bad for all English grammars.
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

These items are good for most English grammars (21/26).
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

![Diagram showing syllable types and token counts]

- 25 types, 976 tokens: a WAY
- 316 types, 12664 tokens: KIT ty
- 14 types, 1480 tokens: UH OH

These items are good for a few English grammars (5/26).
Regular data filter in action

Constraint-based: **OT**, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

How many items should the stress “rule” apply to? \( N = 25 + 316 + 14 = 355 \)
Regular data filter in action

Constraint-based: **OT**, with inflectional knowledge

406 syllable word forms
8 syllable distinctions

How many exceptions are allowed? $355 / \ln(355) = 60$
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

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

How many exceptions are allowed? \(\frac{355}{\ln(355)} = 60\)
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

If this is the dominant pattern, NOT too many exceptions:

\[25 + 14 < 60\]

How many exceptions are allowed? \(355 / \ln(355) = 60\)
Regular data filter in action

**Constraint-based:** **OT**, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

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

How many exceptions are allowed? \( \frac{355}{\ln(355)} = 60 \)
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 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 underlying generative system.
Regular data filter in action

Constraint-based: OT, with inflectional knowledge

- 406 syllable word forms
- 8 syllable distinctions

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