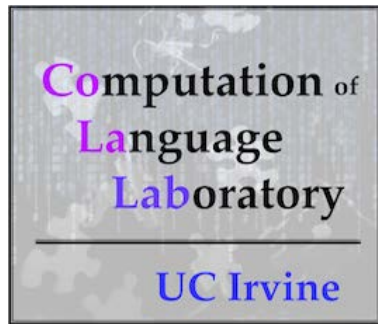


# More learnable than thou?

Testing metrical phonology representations  
with child-directed speech

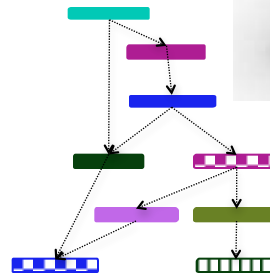
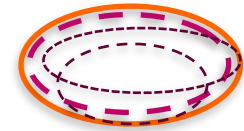
Lisa Pearl

University of California, Irvine



Zephyr Detrano

Tim Ho



February 9, 2014:

40th Annual Meeting of the Berkeley Linguistic Society

University of California, Berkeley

# Grammar variation

A language's grammar is often thought of as a **generative system** speakers use to produce and comprehend the language.

The **generative system can be instantiated in various ways** so that it can **handle any of the world's languages**, and these different ways seem to have something in common.



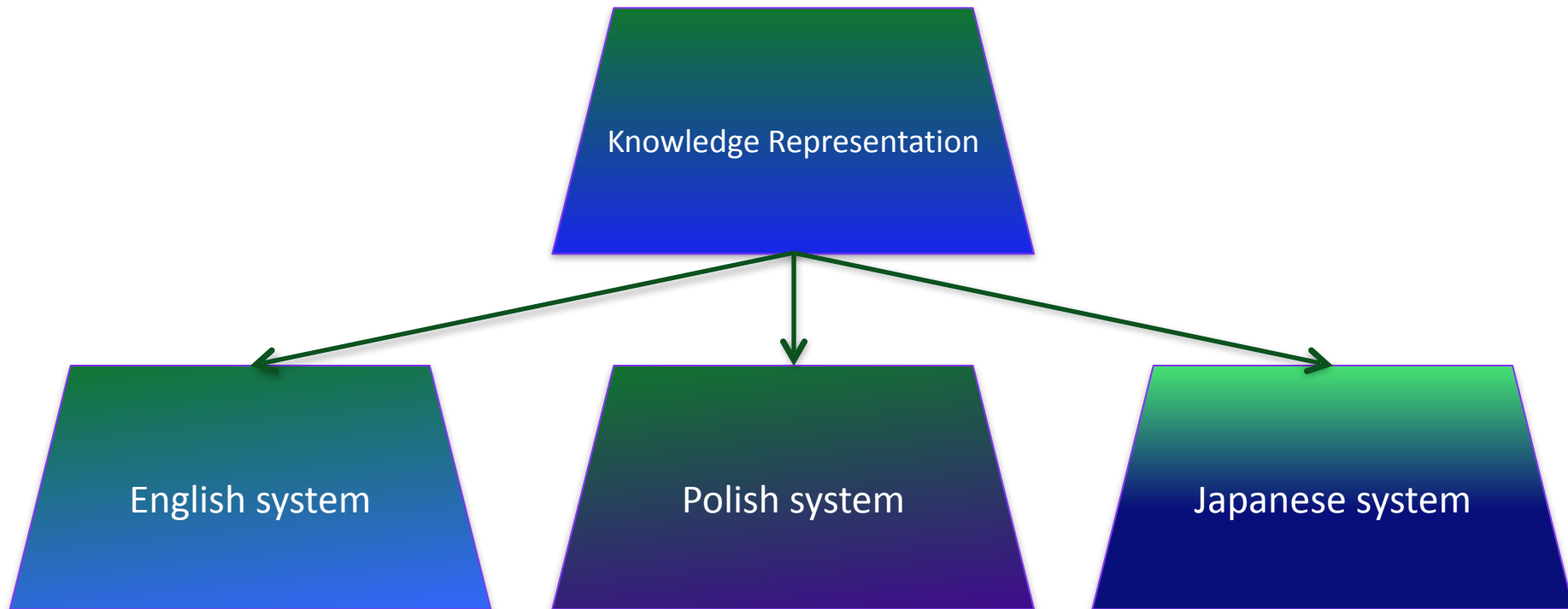
English system

Polish system

Japanese system

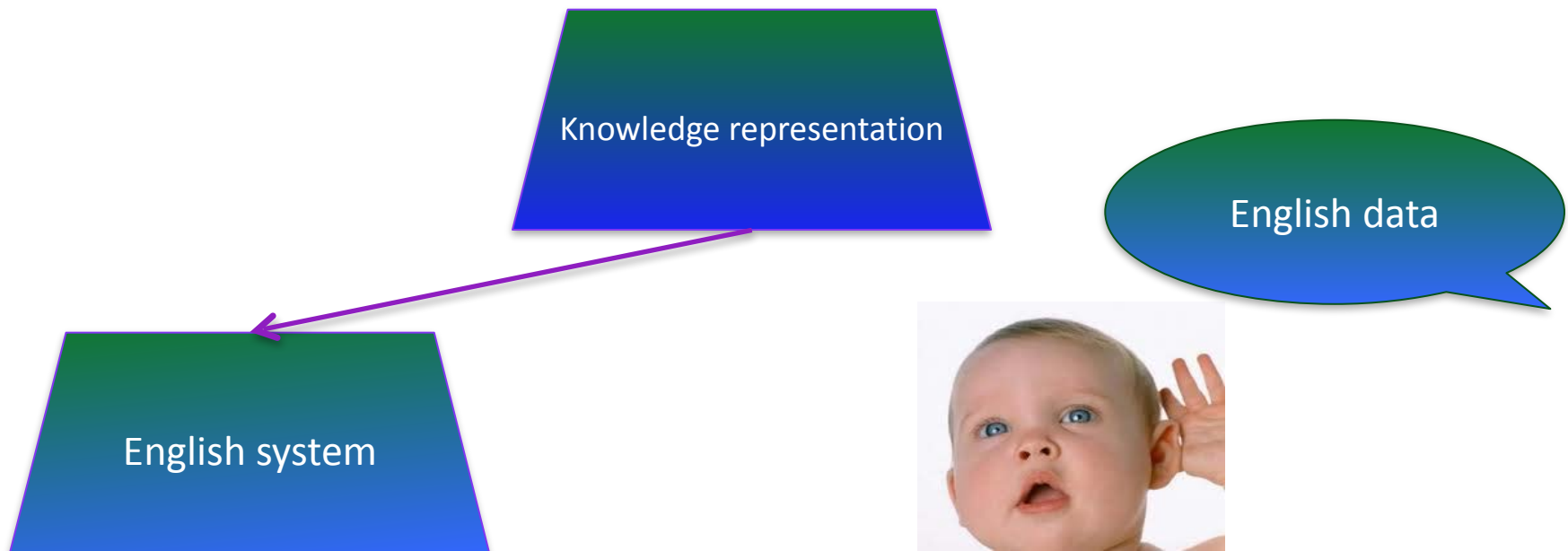
# Grammar instantiations = Knowledge representation

The **knowledge representation** encodes information about **the general form** that grammars for human languages can have.



# Grammar instantiations = Knowledge representation

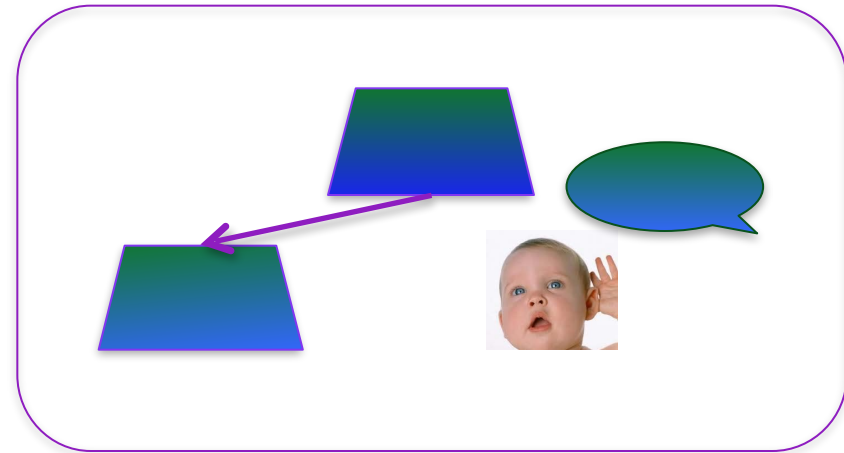
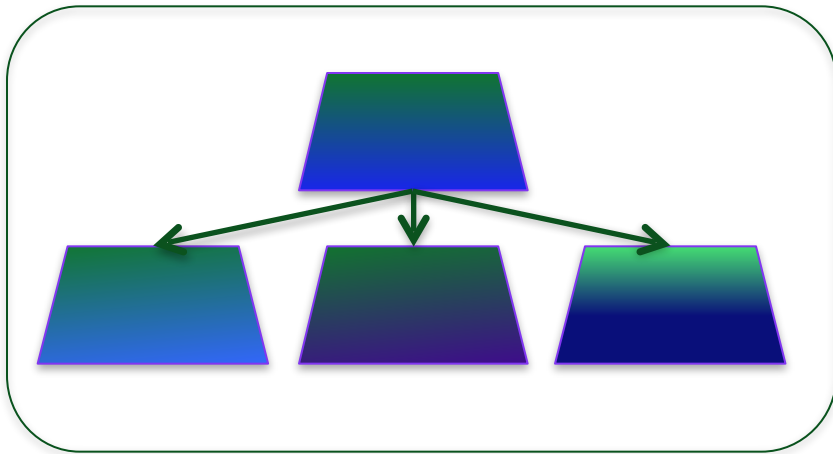
Idea: If the child already knows **the general form that grammars for human languages can have**, all she needs to do is **learn to instantiate her language's grammar appropriately**, based on the input data from her language.



# Knowledge representations

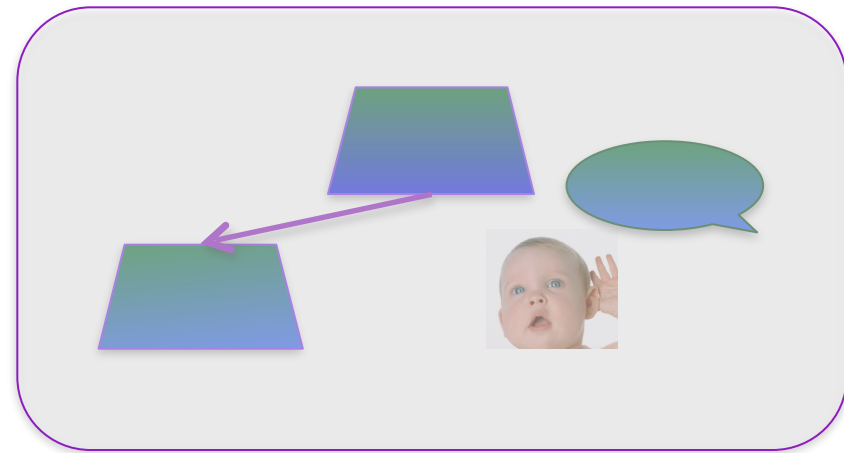
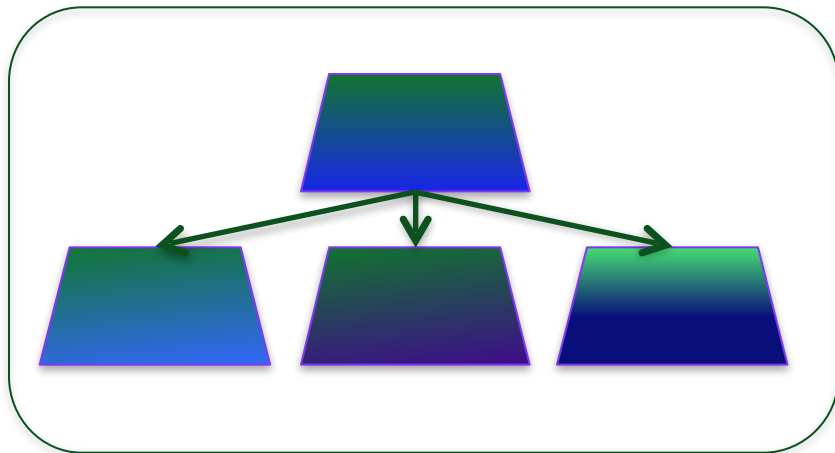
This suggests two checkpoints for any knowledge representation:

- (1) it should explain **constrained cross-linguistic variation**
- (2) its language-specific grammar **should be learnable** from the data children encounter



# Knowledge representations

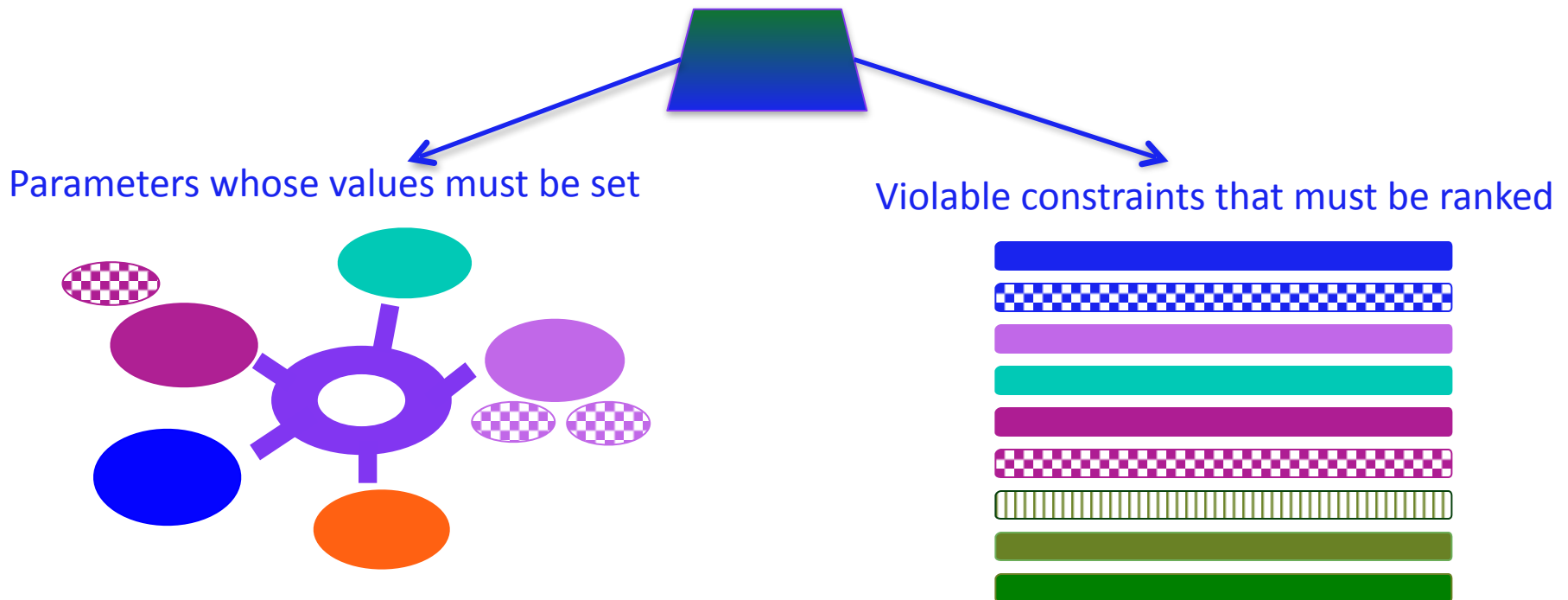
Traditionally, proposals for knowledge representations have focused on optimizing the first checkpoint of accounting for **constrained cross-linguistic variation**, with the (often implicit) assumption that the second checkpoint of **learnability** would be easily satisfied.



# Knowledge representations

## Focus on cross-linguistic variation:

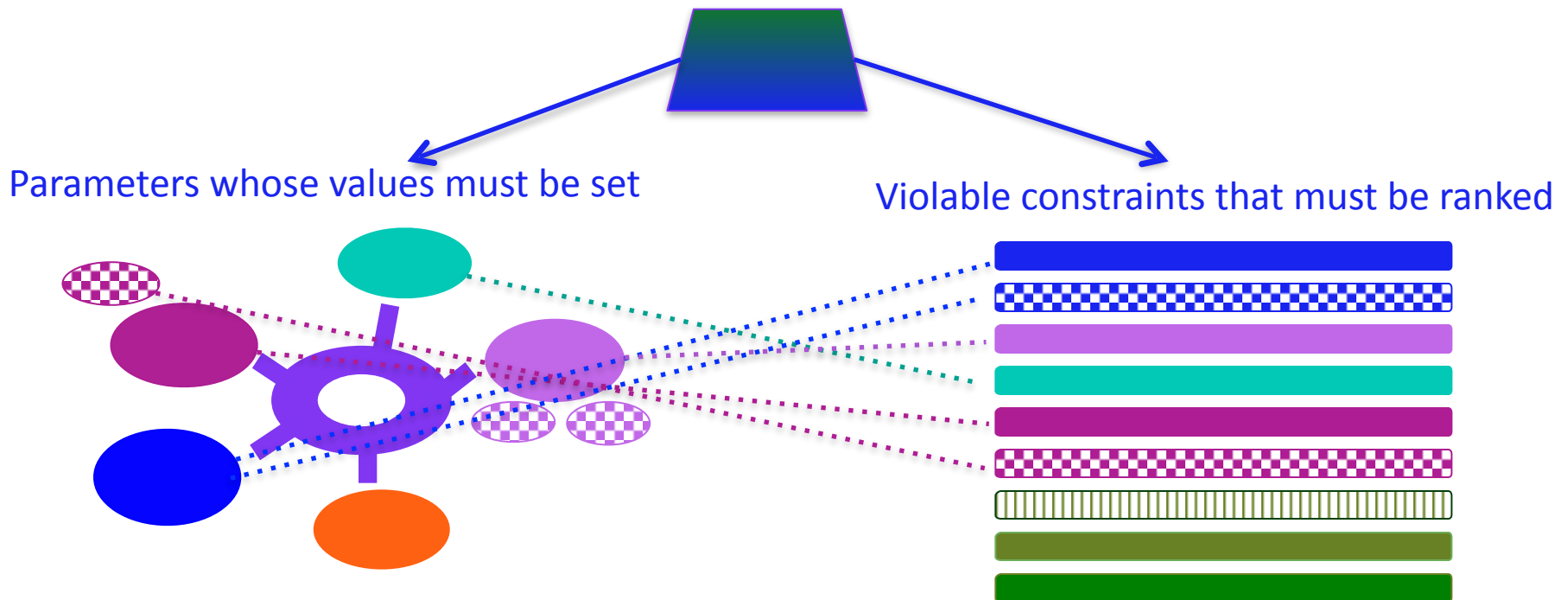
This has led to several knowledge representations for the different aspects of linguistic knowledge. For example, proposals in metrical phonology have included both parametric and constraint-ranking systems.



# Knowledge representations

## Really different:

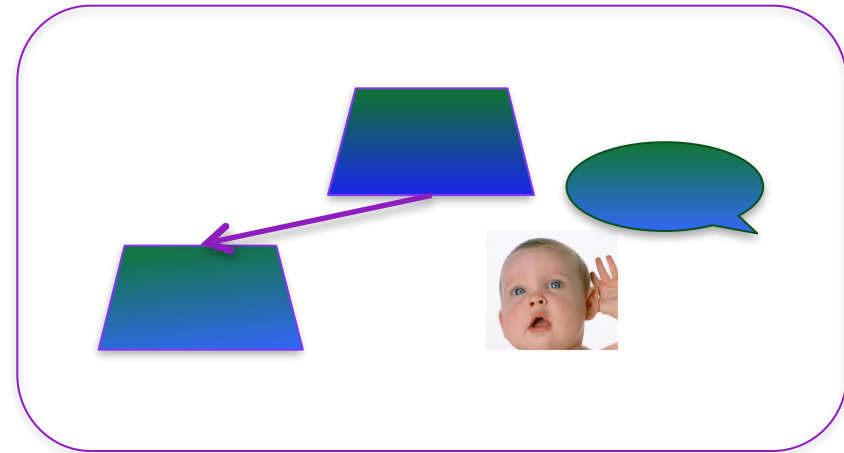
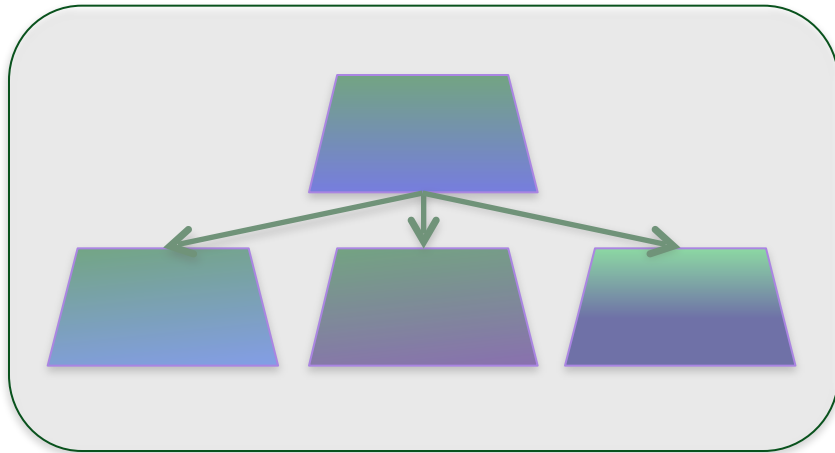
While different knowledge representations will often have aspects in common that they focus on, they are typically not simply notational variants of each other.





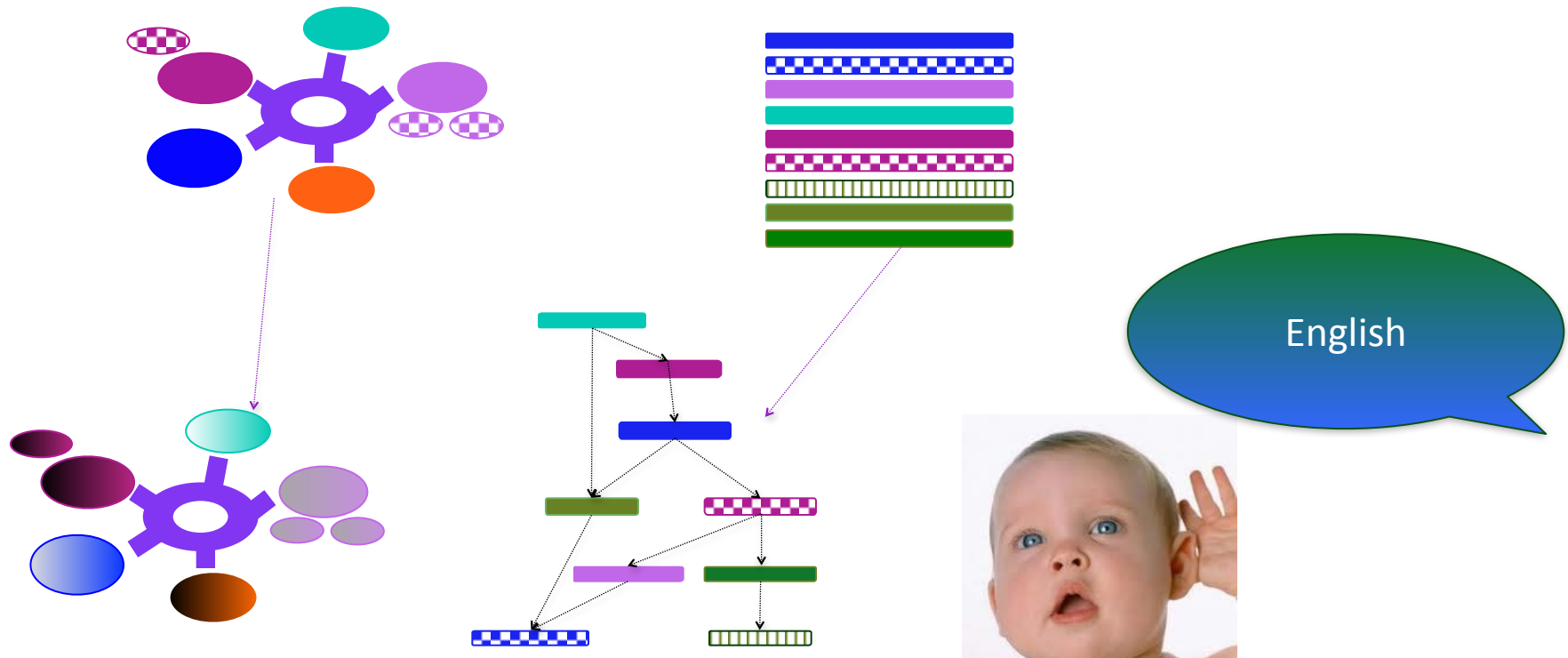
# Knowledge representations

When we have several proposals that satisfy the first criterion of **accounting for constrained cross-linguistic variation**, it seems reasonable to focus on the second criterion of **learnability**.



# Knowledge representations

**Learnability criterion:** Is a language's specific grammar in that knowledge representation **learnable** from the kind of data children of the language encounter?



# Road map

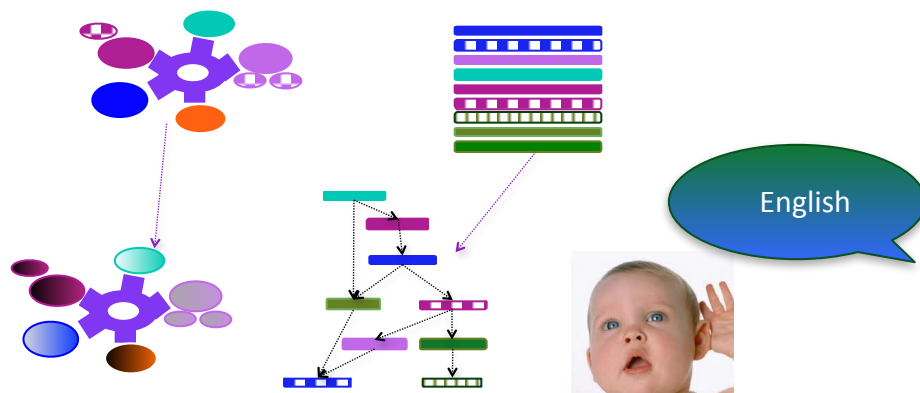
Empirically grounding & quantifying learnability



Case study:

Knowledge representations in metrical phonology

- ❖ Knowledge representation comparison
- ❖ English as a tricky learning scenario
- ❖ Learnability results & implications



# Road map

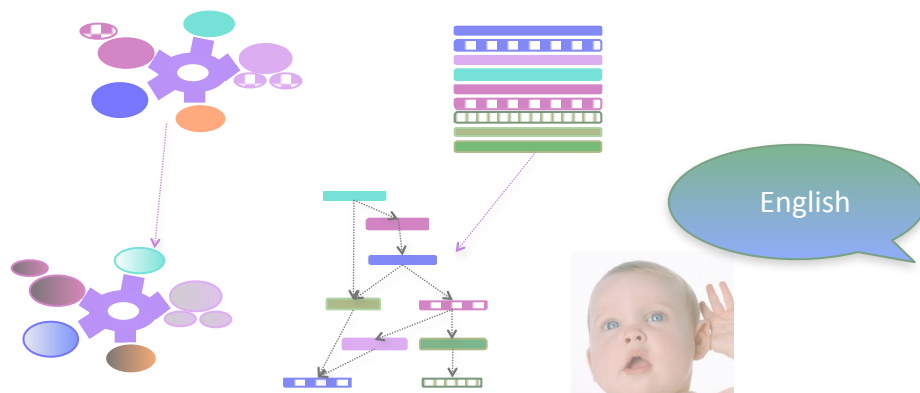
Empirically grounding & quantifying learnability



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

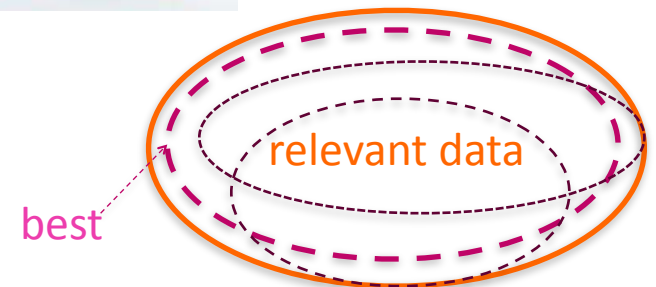
How **easily** does a knowledge representation allow children to learn their specific language's grammar, when given realistic data?

**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 = most useful to have



# 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**.





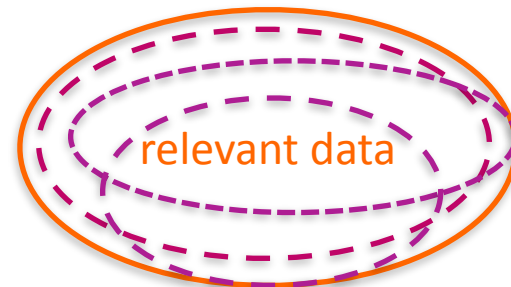
# Quantifying learnability

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.



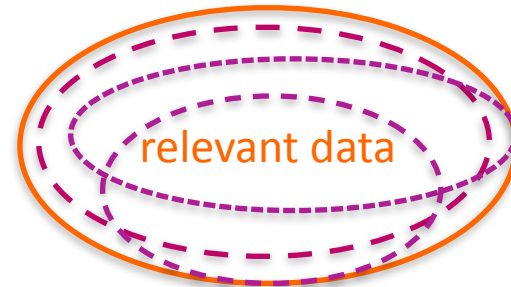
# Quantifying learnability

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 a raw compatibility of 0.70.



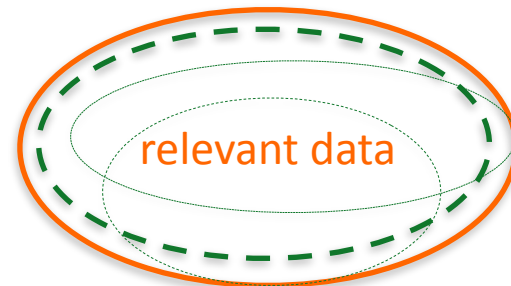
# Quantifying learnability

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 to choose it, given the hypothesis space of grammars.

Example: The best grammar in the knowledge representation has  $\sim 1.00$  relative compatibility.



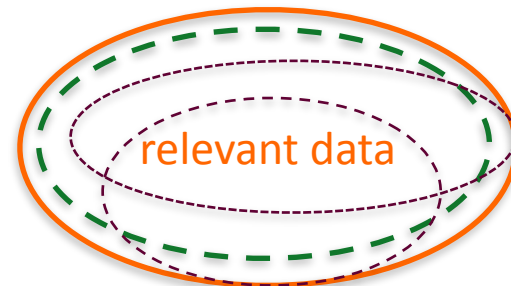
# Quantifying learnability

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  $\approx 1.00$ ) 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 has a learnability potential of 0.70.



# Quantifying learnability

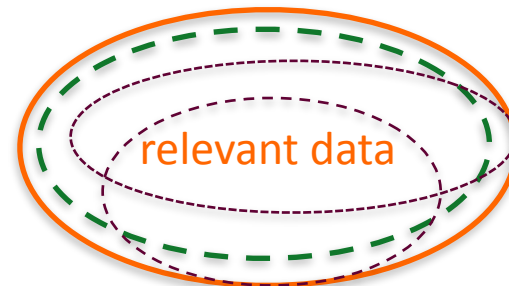
Once we define the **data set the child is learning from**, we can then ask which grammar in the hypothesis space defined by the knowledge representation is **best**, assuming a rational learner that will choose the grammar **compatible with the most data**.



Working assumption:

The **language-specific grammar** should be the best grammar (**relative compatibility  $\approx 1.00$** ) 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.



# Road map

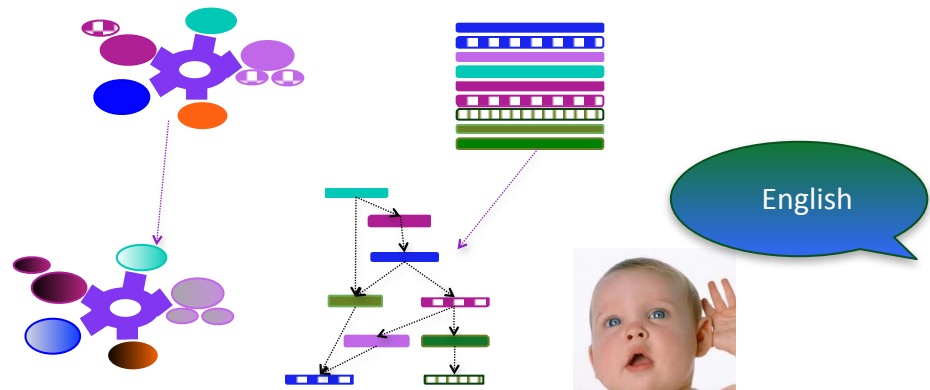
Empirically grounding & quantifying learnability



Case study:

Knowledge representations in metrical phonology

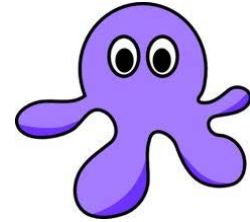
- ❖ Knowledge representation comparison
- ❖ English as a tricky learning scenario
- ❖ Learnability results & implications



# Case study: A generative system of metrical phonology

Observable data: **stress contour**

**O**Ctopus



Underlying representation determined by grammar?

# Case study: A generative system of metrical phonology



Observable data: **stress contour**

**OC**topus

Underlying representation determined by grammar?

( H L ) H  
VC V VC  
OC to pus

( S S ) S  
VC V VC  
OC to pus

( H L L )  
VC V VC  
OC to pus

( S S S )  
VC V VC  
OC to pus

Divide word into syllables,  
look only at syllable rimes

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

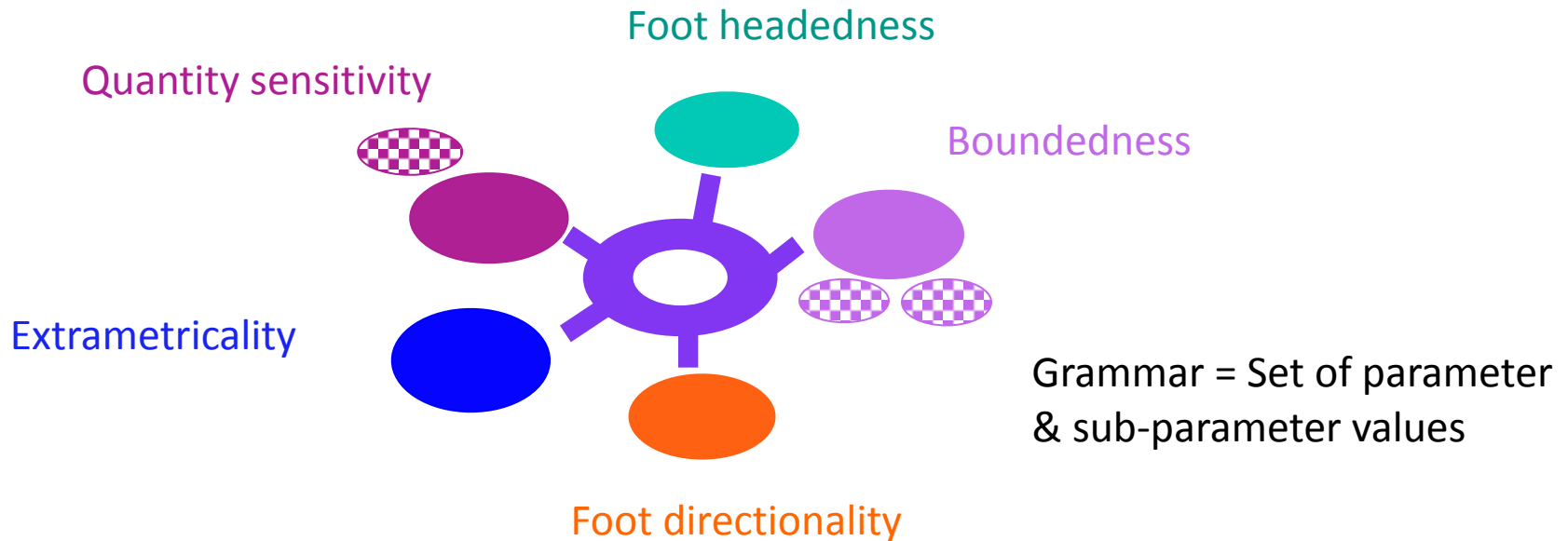
**HV**: Halle & Vergnaud 1987, Dresner 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars

Correct grammar **builds**  
compatible contour

**O**ctopus



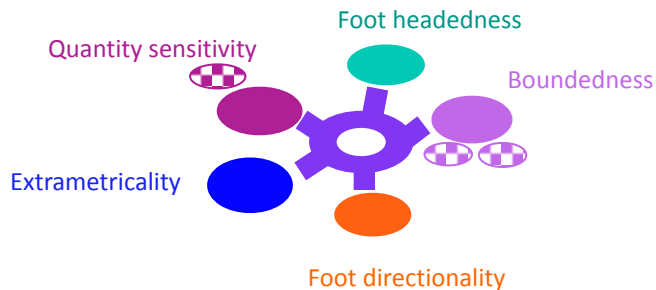
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Parameter values used:

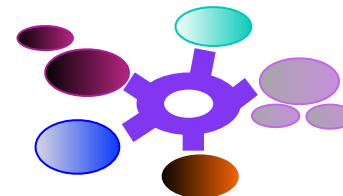
QS-VC-H, Em-Rt, FtDir-Rt, B-2-Syl, FtHd-Left

Correct grammar builds compatible contour

OCTopus

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

(H L) H  
OC to pus



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

# Three knowledge representations

## Parametric systems

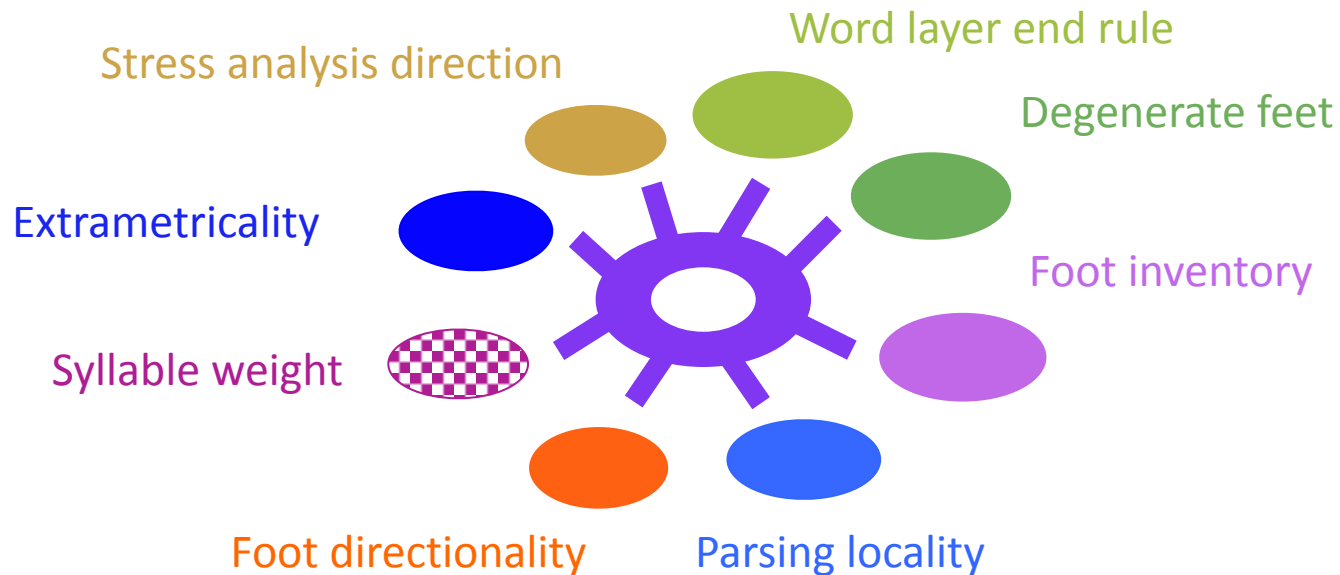
**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Correct grammar **builds**  
compatible contour

**O**ctopus



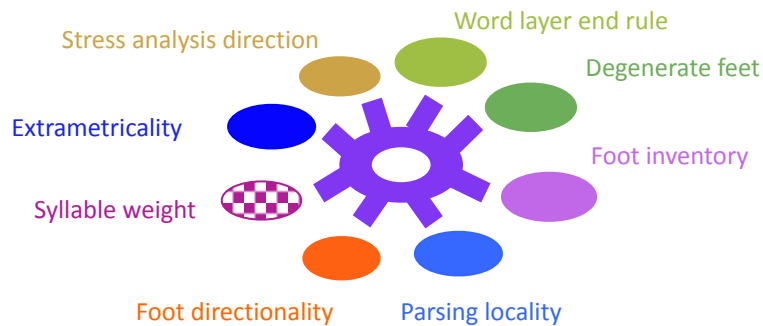
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Parameter values used:

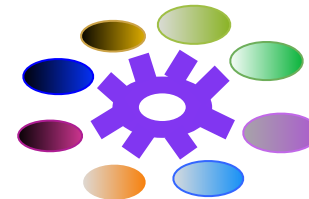
Bot, Em-RtCons, VC-H, FtDir-Rt,  
PL-Strong, MorTro, DF-Strong, WLER-Rt

Correct grammar builds compatible contour

OCTopus

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

( H )( L L )  
OC TÓ pus



...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

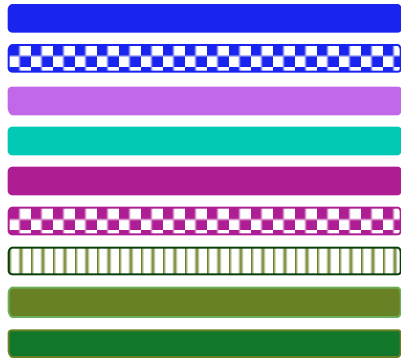


# Three knowledge representations

## Constraint-ranking systems

**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.

Best candidate for the correct grammar has a compatible contour

Octopus

(OC to) pus

oc (TO pus)

(oc TO) pus

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

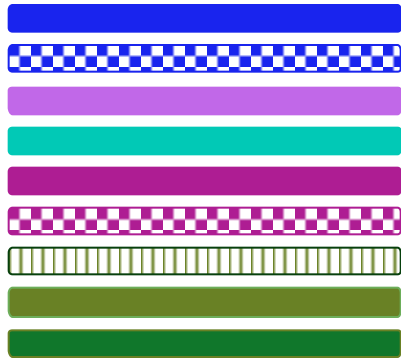
# Three knowledge representations

## Constraint-ranking systems

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

9 violable constraints

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



Grammar = ranked ordering of all constraints

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

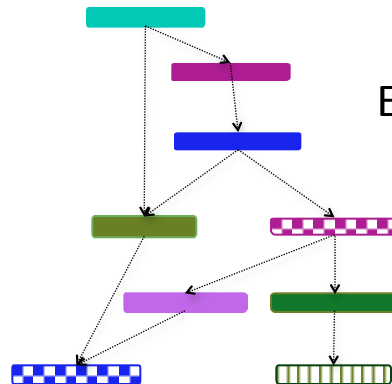
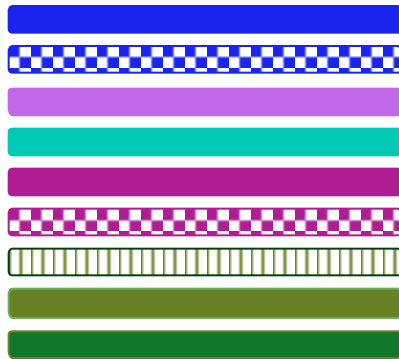
9 violable constraints

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

Best candidate for the correct grammar has a compatible contour

Octopus

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



English grammar

# Three knowledge representations

## Constraint-ranking systems

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

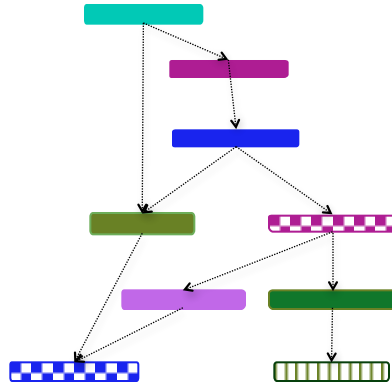
9 violable constraints

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

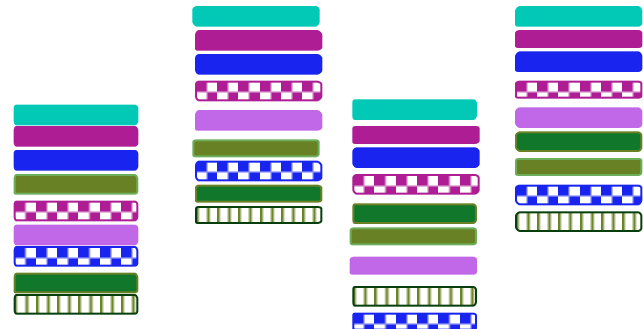
Best candidate for the correct grammar has a compatible contour

OCTopus

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

# 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



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)

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



A sample grammar that is a version of the English “grammar”:



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)

# 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

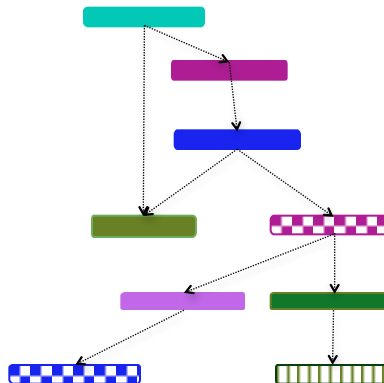


A sample grammar that is a version of the English “grammar”:



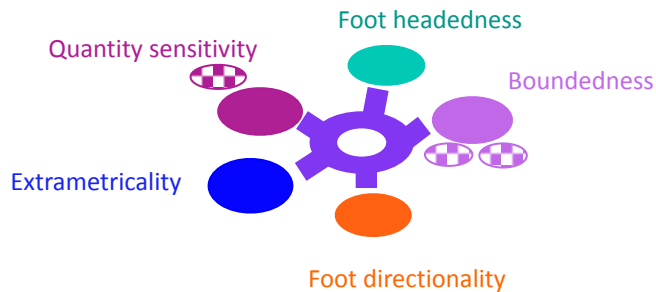
Best candidate for the correct grammar has a compatible contour

OCTopus



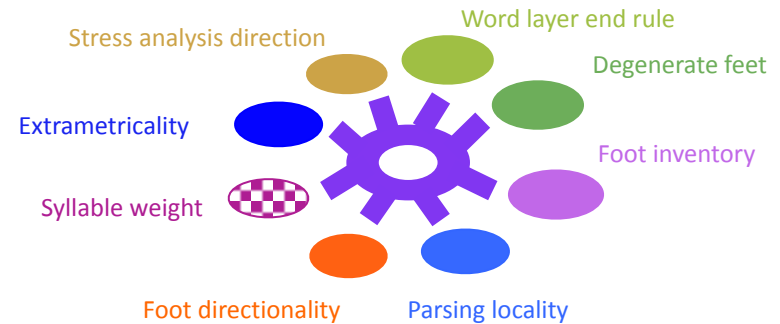
English “grammar”

# Knowledge representation comparison



**HV:** 5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



**Hayes:** 8 parameters

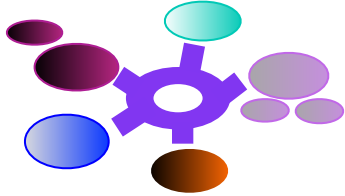
Hypothesis space: 768 grammars



**OT:** 9 violable constraints

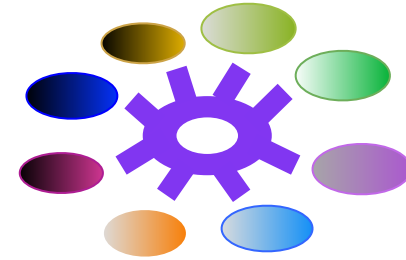
Hypothesis space: 362,880 grammars

# English instantiations



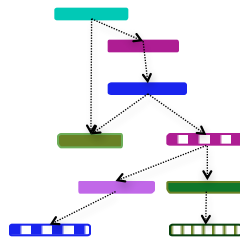
**HV:** 5 parameters & 4 sub-parameters

Hypothesis space: 156 grammars



**Hayes:** 8 parameters

Hypothesis space: 768 grammars



**OT:** 9 violable constraints

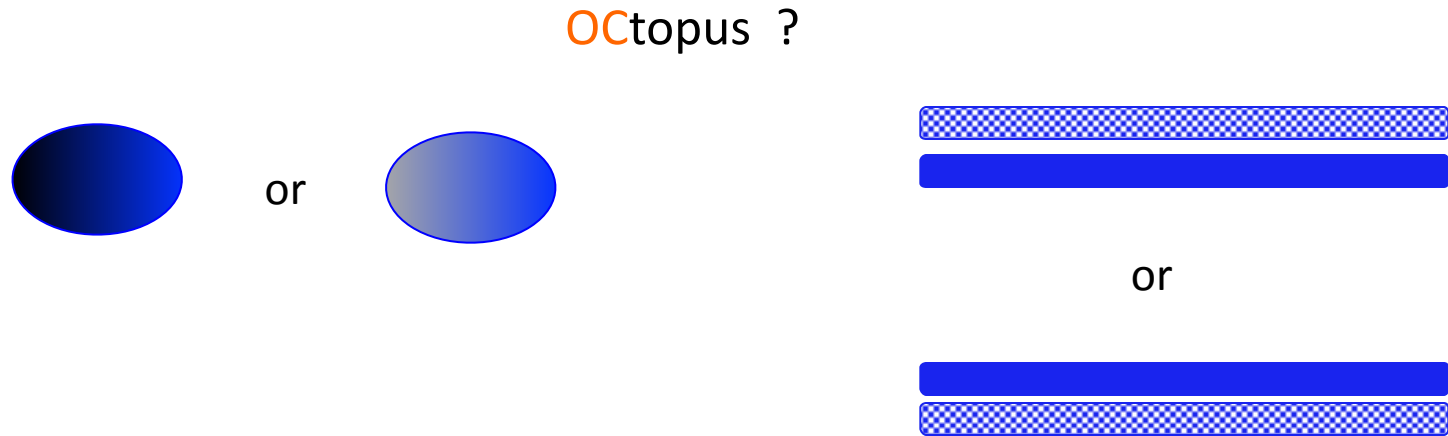
Hypothesis space: 362,880 grammars

(English = 26 grammars)



# 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 (**poverty of the stimulus** = the data are compatible with many hypotheses).

# 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.

**EAR**ly

**EAR**lier

**PRE**tty

**PRE**ttiest

sen**S**ation

sen**S**ational

sen**S**ationally

# 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

CONduct

DEsert

SUspect

## VERBS

conDUCT

deSERT

suSPECT

# 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

V vv

a WAY

v VV

UH OH

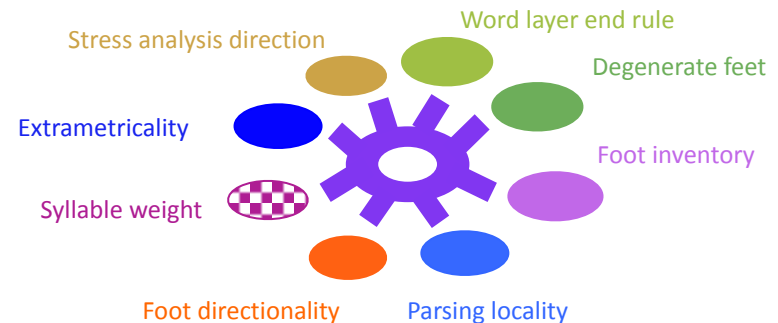
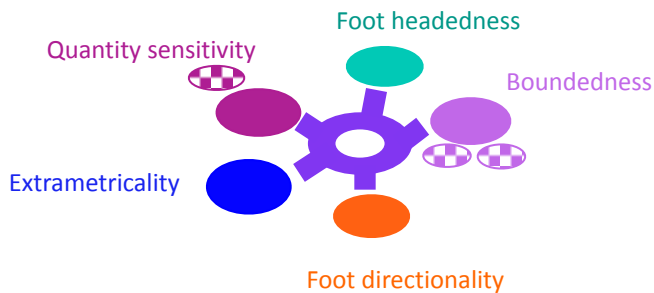
V VV

# 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...

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**.

Syllabic word form: V VV

Select  
one of these...

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



# 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

V vv

a WAY

v VV

UH OH

V VV



# 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

V vv

a WAY

v VV

UH OH

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!



# 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  $\approx 1.00$ ) 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  $\approx 1.00$ ) can account for

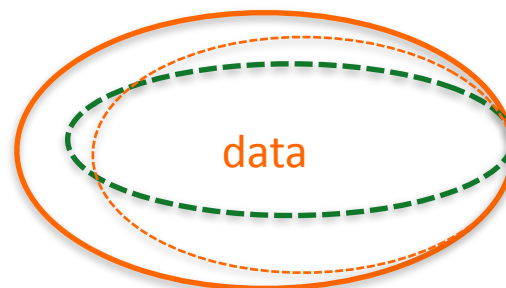
## Raw compatibility of best grammar

**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.



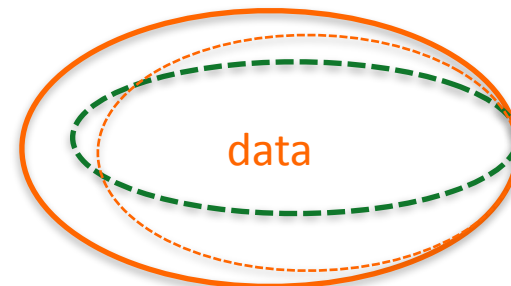
## Raw compatibility of best grammar

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



# 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

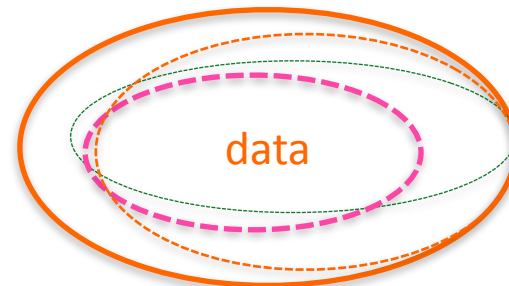
## Raw compatibility of English grammar

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



# 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)

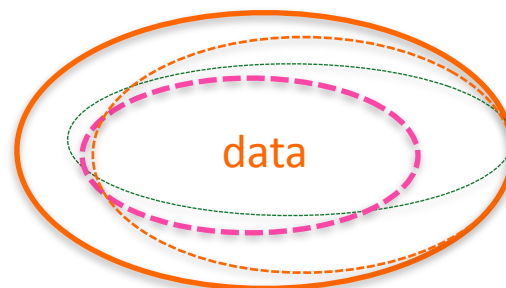
Significantly less than the best grammar

## 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)

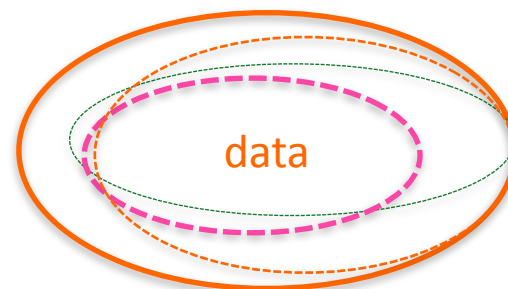
Significantly less than the best grammar

## 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

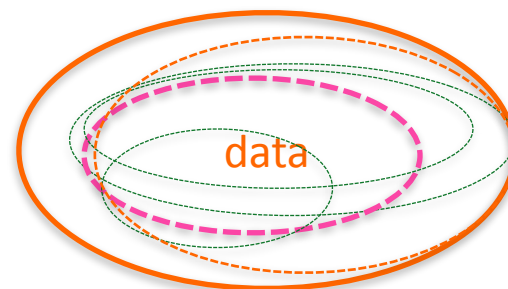
## Relative compatibility of English grammar

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.



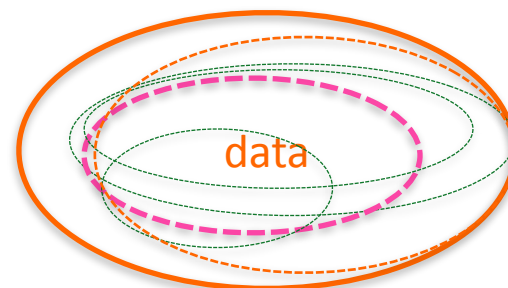
## Relative compatibility of English grammar

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

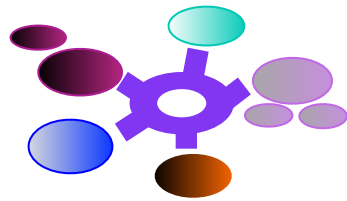
Better than many...but many are still better



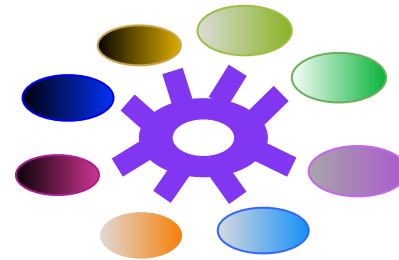
# Learning English metrical phonology

Interim conclusion:

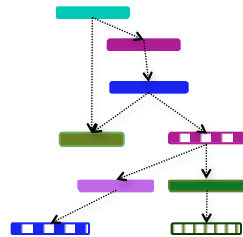
**Learnability issues exist** for the English grammar in all three knowledge representations.



Parametric: HV

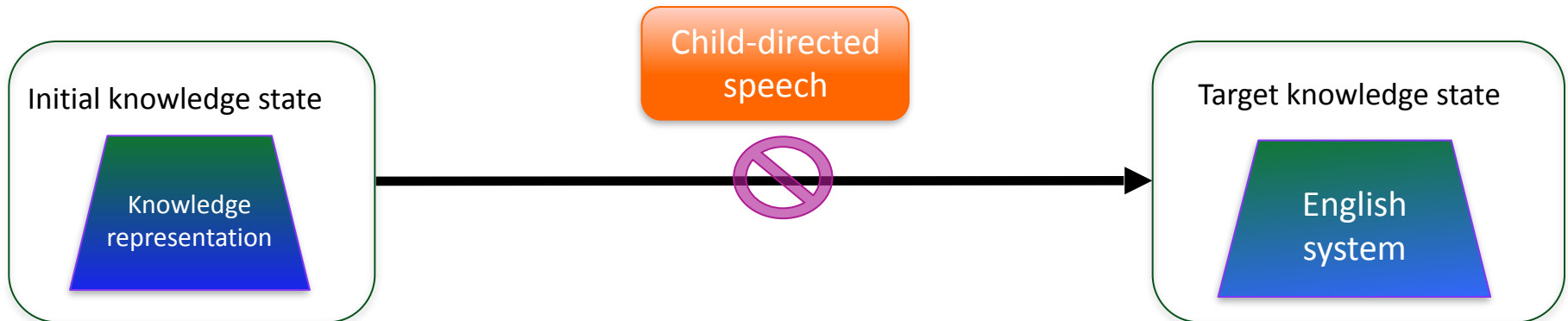


Parametric: Hayes

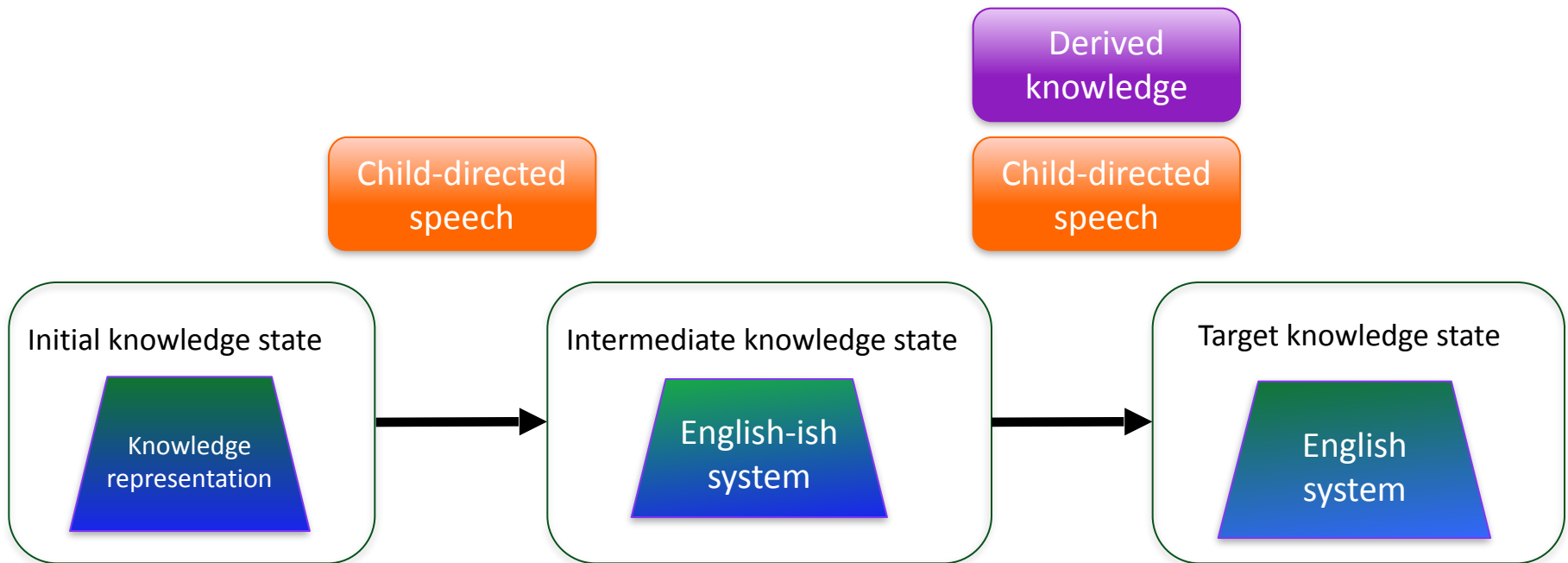


Constraint-based: OT

# 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.

**EAR**ly

**EAR**lier

**PRE**tty

**PRE**ttiest

sen**S**ation

sen**S**ational

sen**S**ationally

# 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

EARlier

PREtty

PREttiest

senSAtion

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

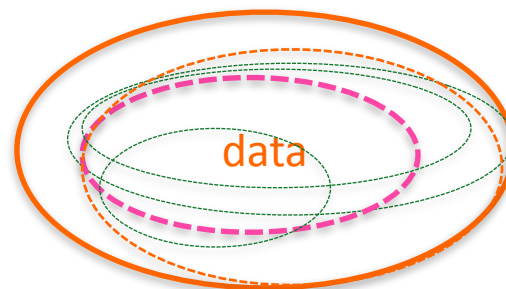
## Relative compatibility of English grammar

HV: 0.712 by types (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:

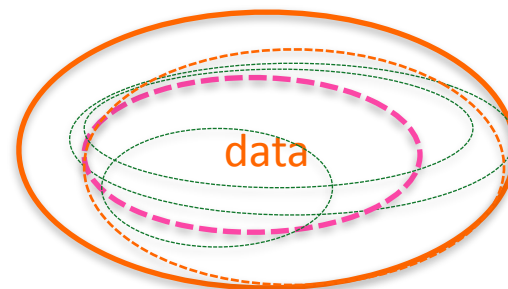
There remain many other grammars in the hypothesis space that are more compatible with the data, even though the learner knows inflectional morphology is stressless. Even if children aren't optimal learners, it would be easier to pick one of these other more compatible grammars.



## 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

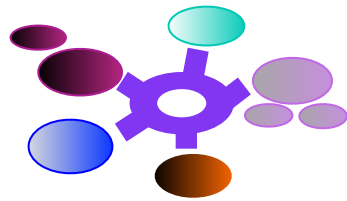
Better than many...but many are still better



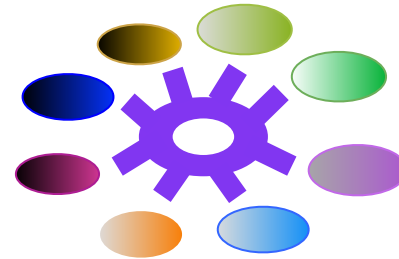
# Learning English metrical phonology

Continuing conclusion:

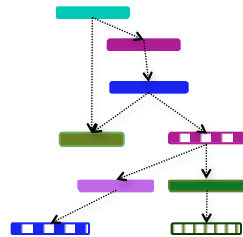
The same learnability issues persist for the English grammar in all three knowledge representations, even when the learner has some knowledge of the interactions between morphology and metrical phonology.



Parametric: HV

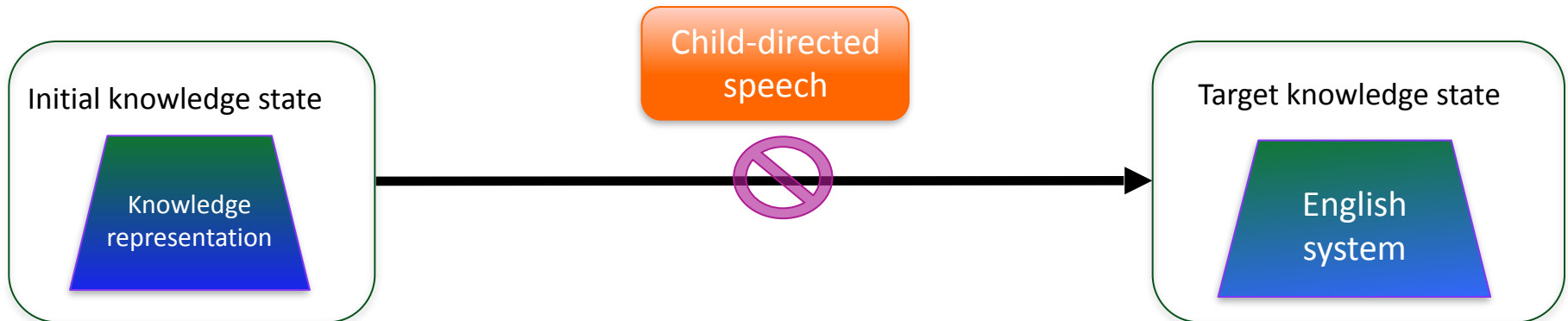


Parametric: Hayes

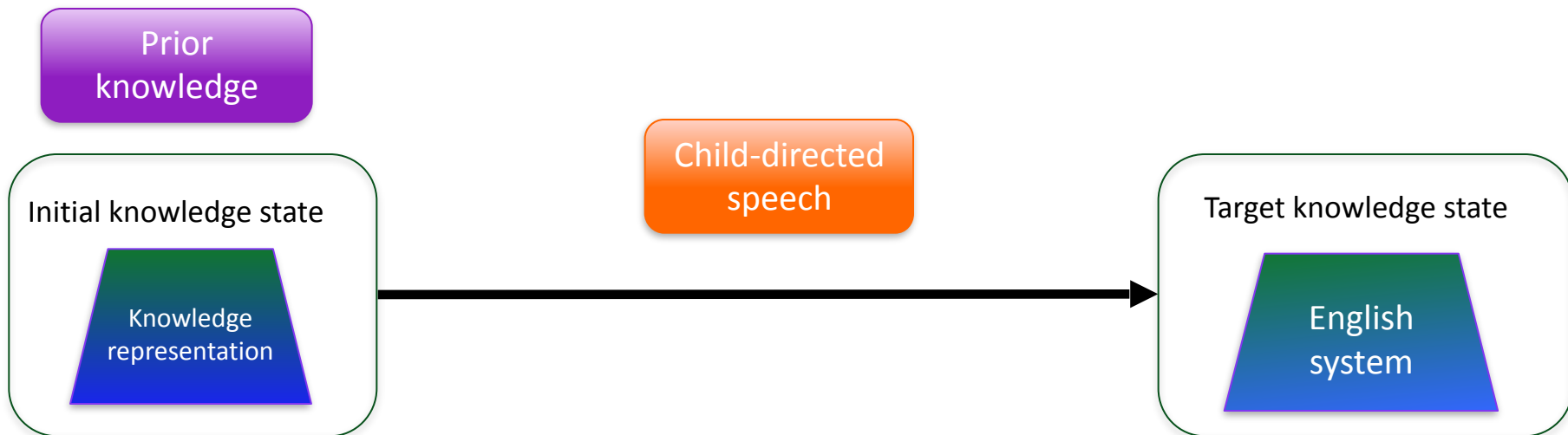


Constraint-based: OT

# The learnability problem: Another option



# The learnability problem: Another option



The initial state of the learner includes **prior knowledge** that helps the learner learn in a more sophisticated way.

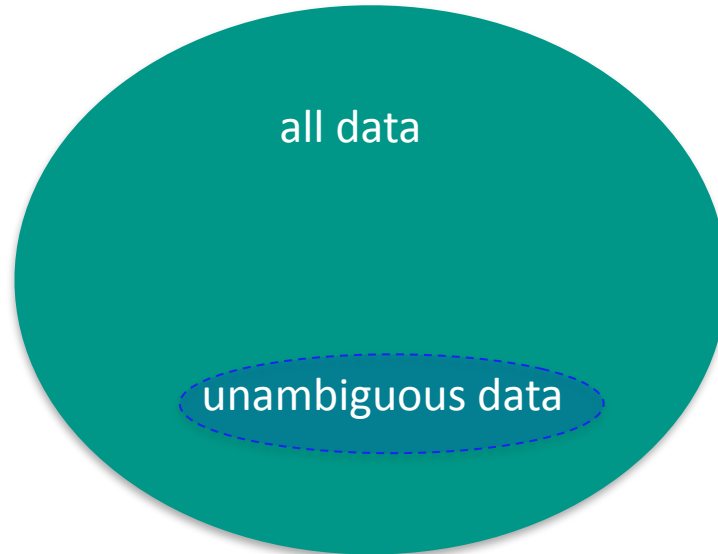
# One type of useful knowledge: Unambiguous data filter

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



# One type of useful knowledge: Unambiguous data filter

**Updated** working assumption: The learner will try to learn a grammar that can account for all the **unambiguous** data encountered (Fodor 1998, Pearl 2008).



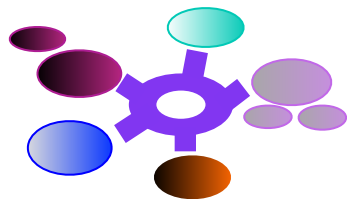
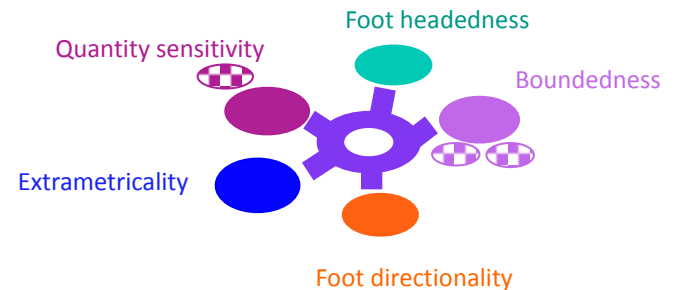
This is a small subset of the available data  
which can be viewed as **maximally informative**.



# One type of useful knowledge: Unambiguous data filter

**Updated** working assumption: The learner will try to learn a grammar that can account for all the **unambiguous** data encountered (Fodor 1998, Pearl 2008).

Pearl 2008: This type of bias will allow the **HV English grammar** to be learned from English data.

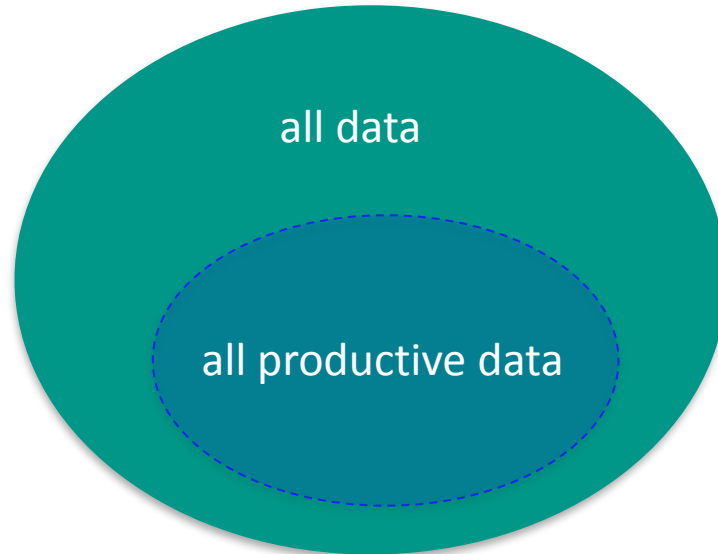


Selectively learning  
from the input



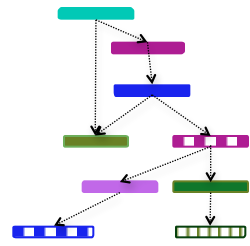
# Another type of useful knowledge: Productive data filter

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



# 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).



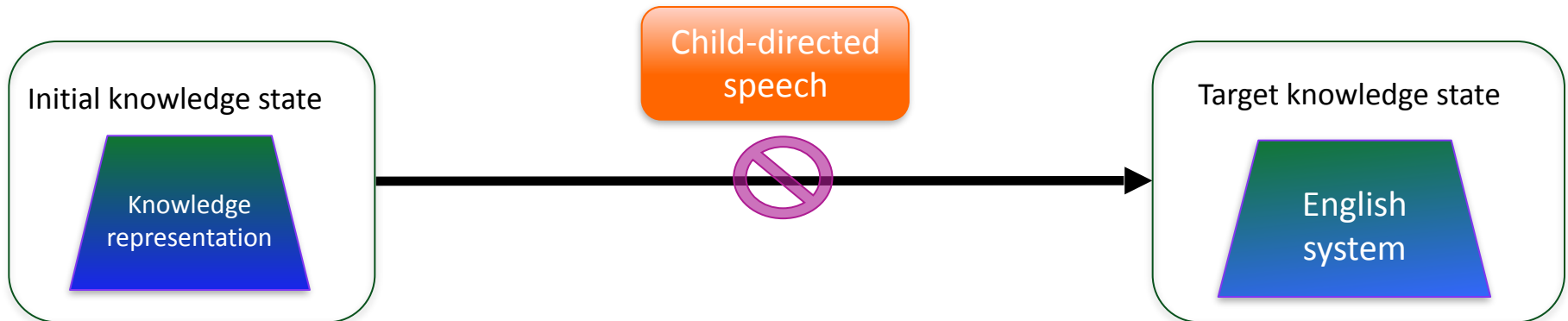
✓  
COO kie  
DA ddy  
FU nny

Selectively learning from the input

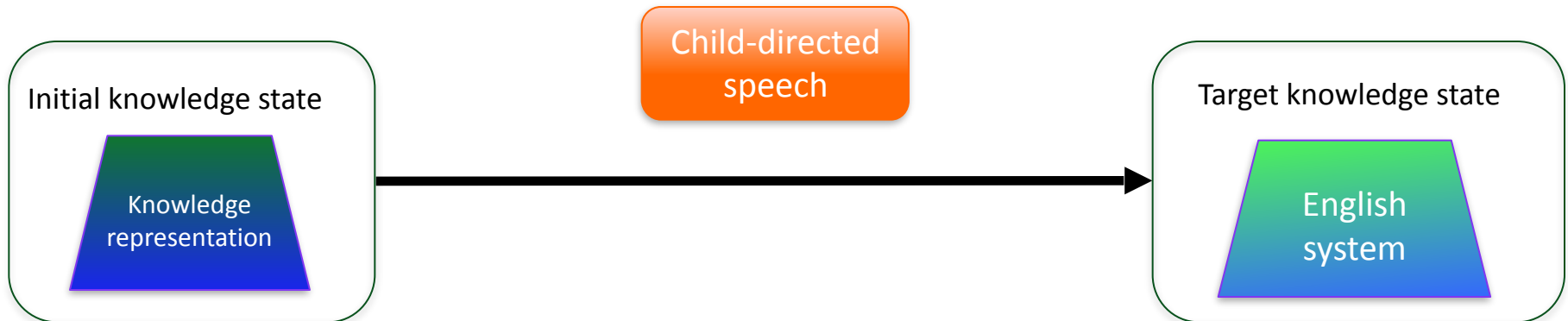


✗  
a WAY  
be LOW  
to DAY

# The learnability problem: A third option



# The learnability problem: A third option

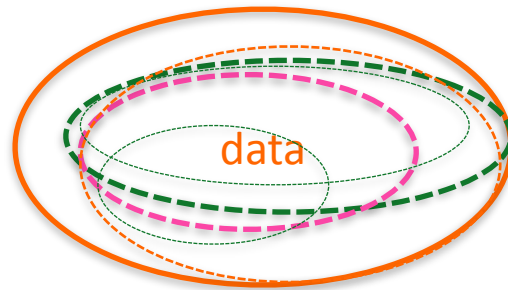


The **target grammar is different** than we think, and we should update our definition of the English grammar.

# What should the target grammar be?

English  
system

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

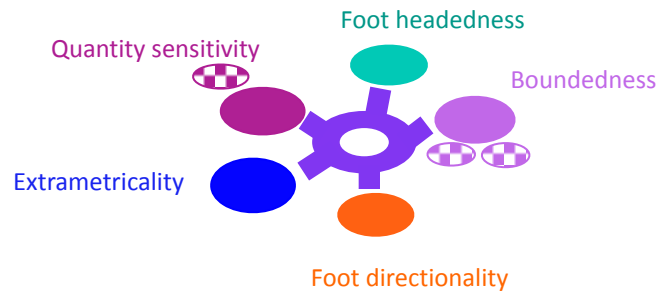


# What should the target grammar be?

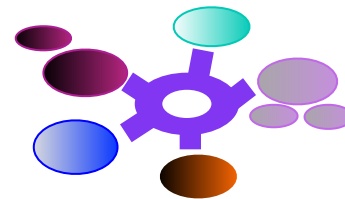
English  
system

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

## Parametric: HV



## English grammar

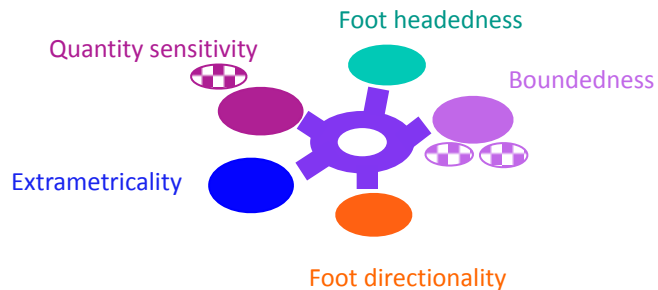


# What should the target grammar be?

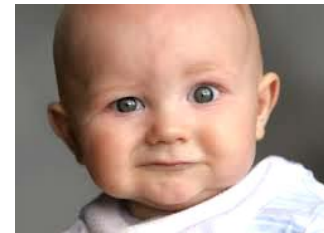
English  
system

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

## Parametric: HV



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.



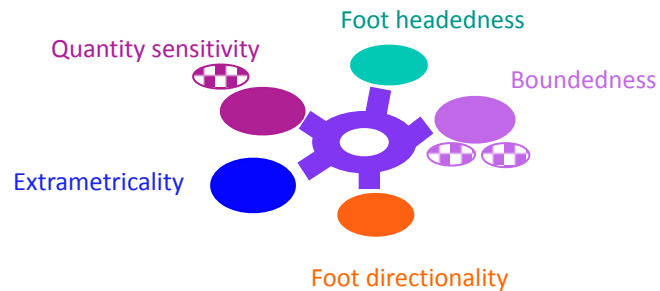


# What should the target grammar be?

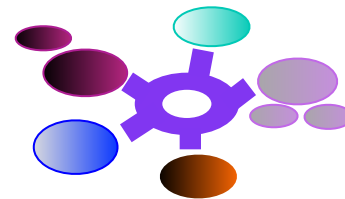
English  
system

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

## Parametric: HV



## English grammar



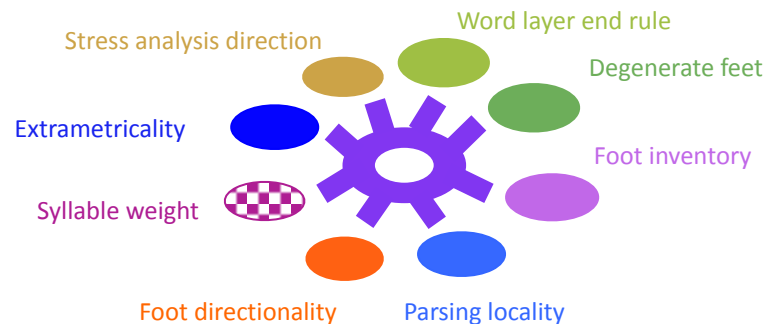
Upshot: Unclear for the HV knowledge representation that the learning problem can be fixed by simply switching one parameter value here or there.

# What should the target grammar be?

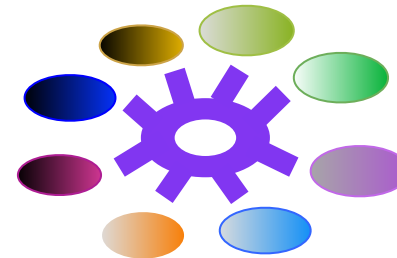
English  
system

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

## Parametric: Hayes



## English grammar

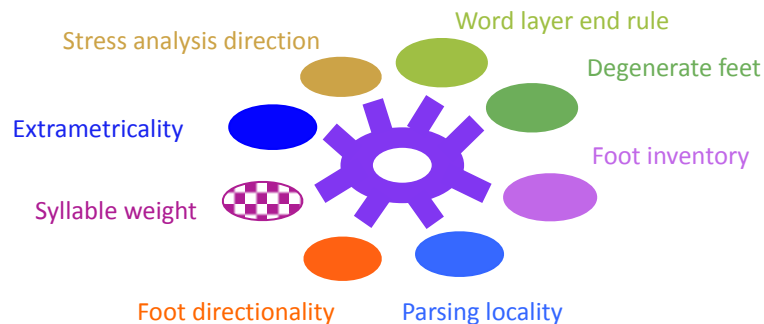


# What should the target grammar be?

English  
system

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

## Parametric: Hayes



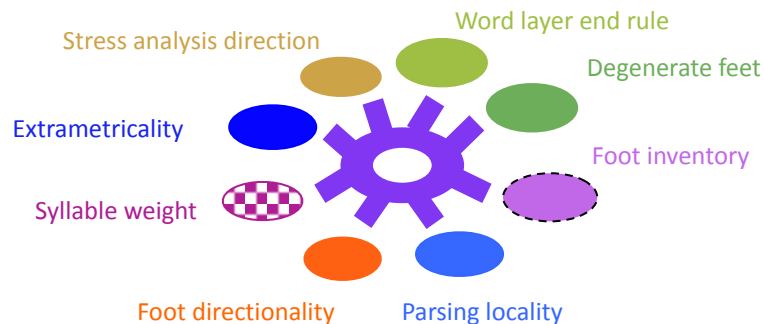
If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.

# What should the target grammar be?

English  
system

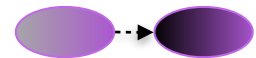
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## 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.



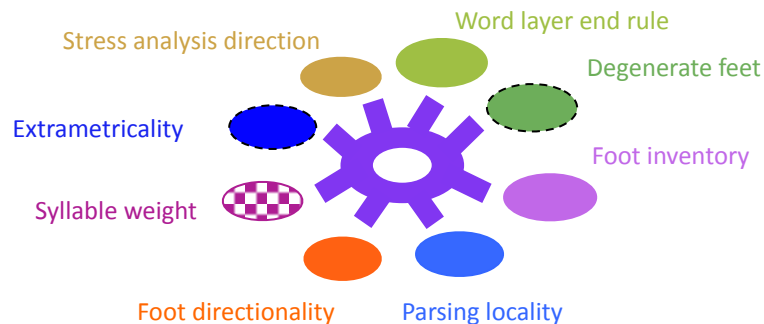
**Relative compatibility = 0.910**

# What should the target grammar be?

English  
system

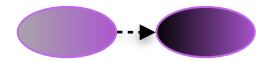
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## 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.



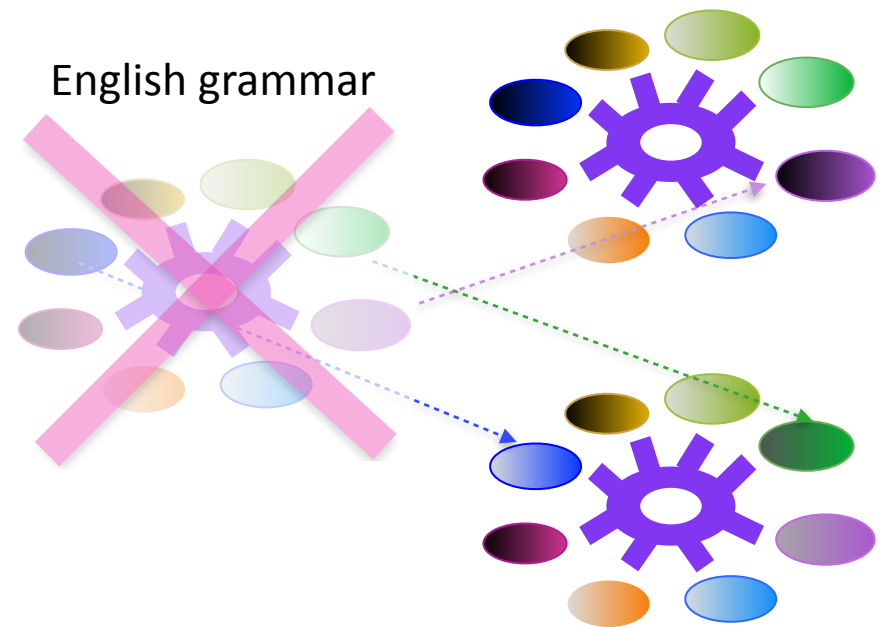
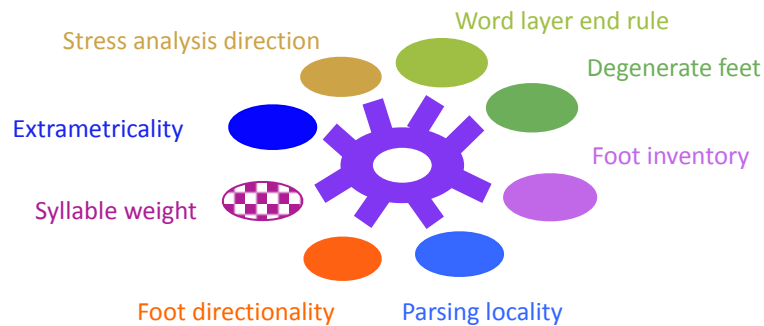
**Relative compatibility = 0.923**

# 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



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?

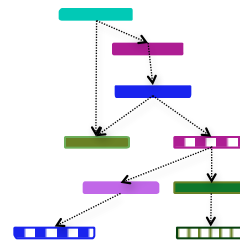
English  
system

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



English grammar obeys this ordering



# What should the target grammar be?

English  
system

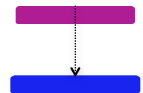
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## Constraint-based: OT



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

Weight-to-Stress VV >> Nonfinality

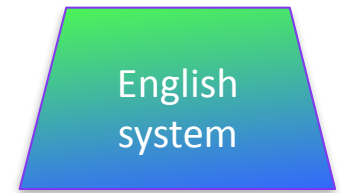


(Effect: Prefer long syllables to be stressed, even if they're at the right edge of the word)

BA**by** ← problematic



# What should the target grammar be?

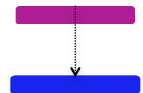


One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

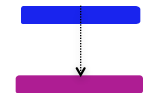
## Constraint-based: OT



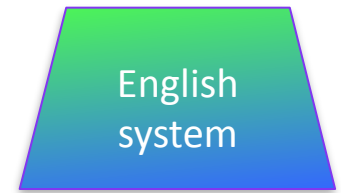
If we examine the grammars with high compatibility, it turns out that there is one constraint ordering the English grammar uses which none of these grammars use.



Implication: This ranking may be better for English data.



# What should the target grammar be?

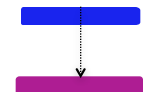
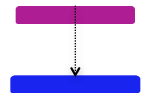


One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## Constraint-based: OT



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



Relative compatibility = 0.988

# What should the target grammar be?

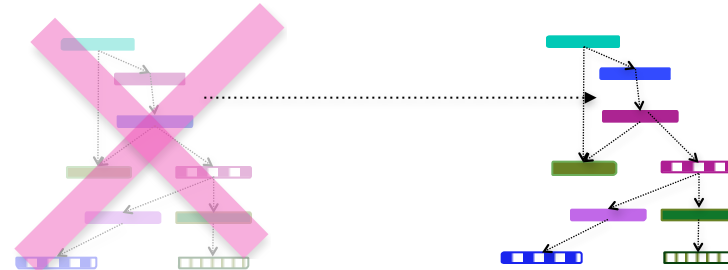
English  
system

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



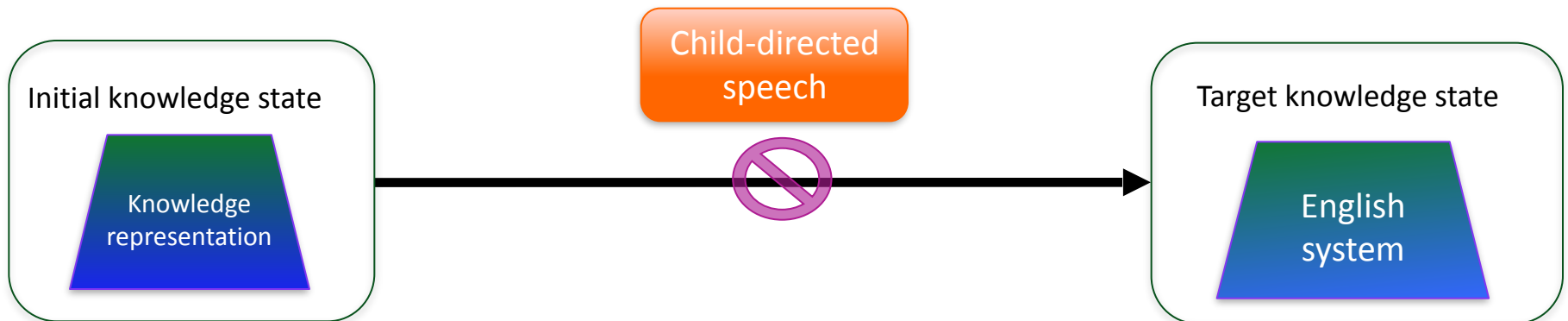
English grammar obeys this ordering



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

# Learnability implication recap

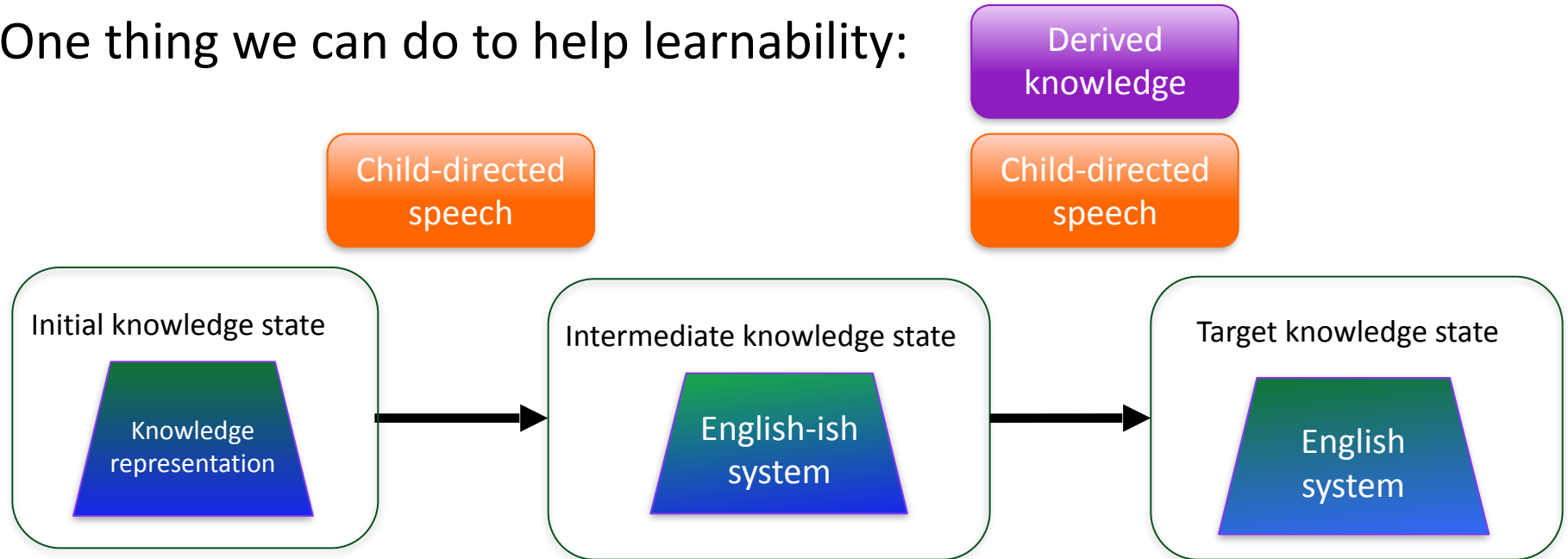
Since learnability issues exist for all three knowledge representations, something else must be going on.



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:

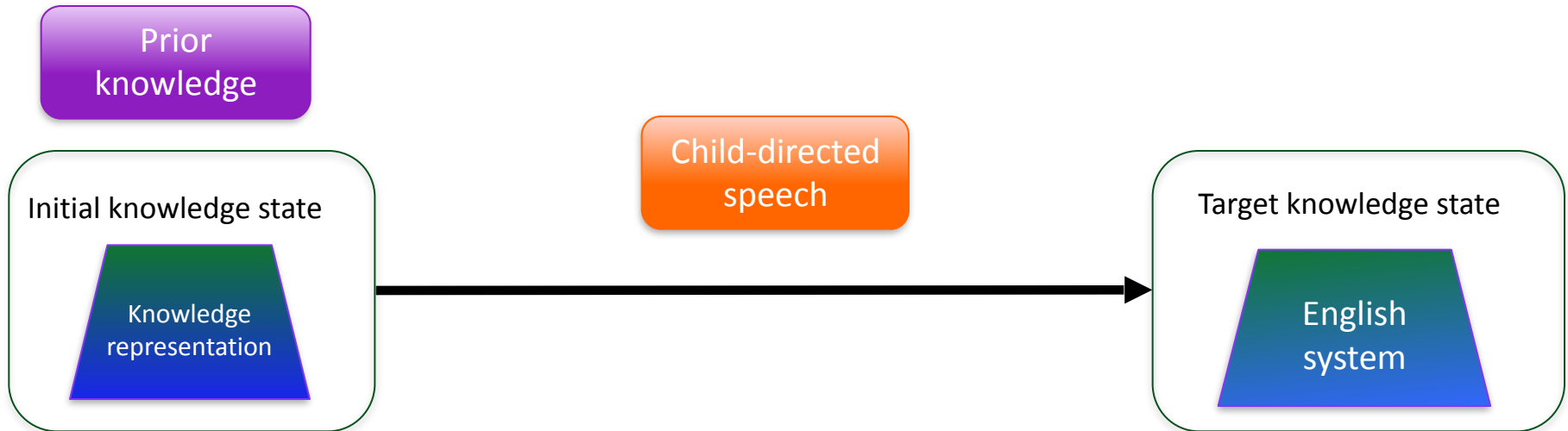


Assume there is at least one transitory state in learning that the learner reaches and then leaves **once additional knowledge is acquired**.

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

# Learnability implication recap

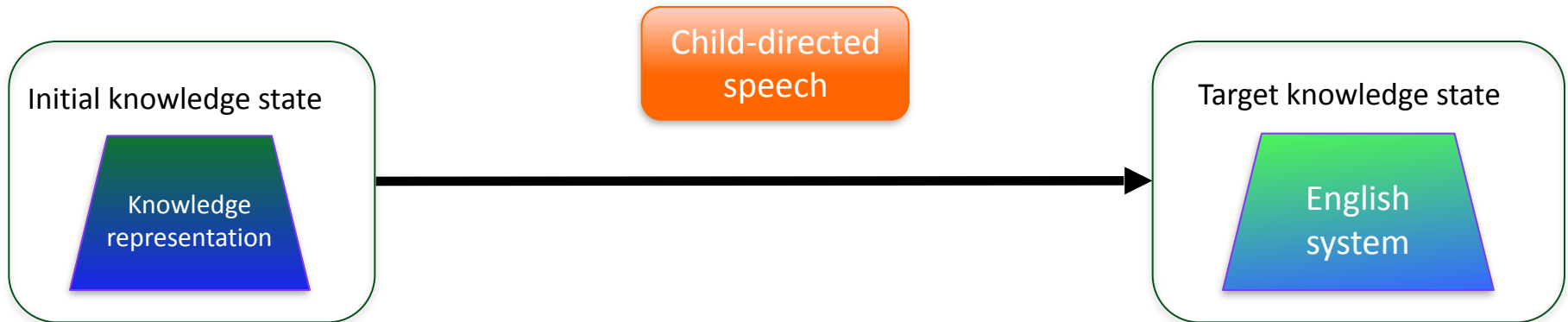
Another thing we can do to help learnability:



Add in **prior knowledge** that helps the child perceive the input in a helpful way. Two potentially useful learning biases involve learning a grammar that accounts for a subset of the available data, rather than all of it. One parametric system's English grammar (**HV**) has been shown to be learnable this way.

# Learnability implication recap

A third thing we can do to help learnability:



Update our ideas about the **target grammar** slightly: alter certain parameter values or constraint rankings. This could allow the parametric **Hayes** and constraint-ranking **OT** English grammars to be easily learnable from English data.

Big picture



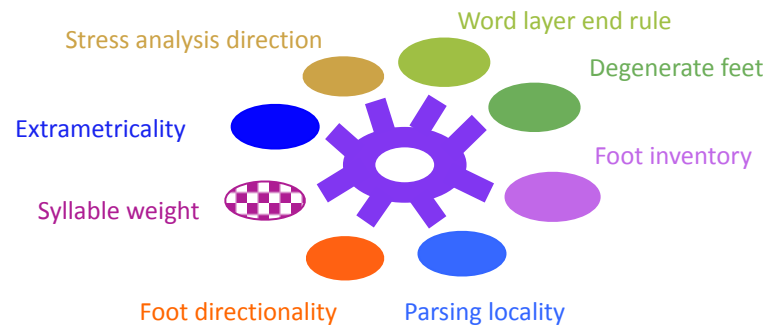
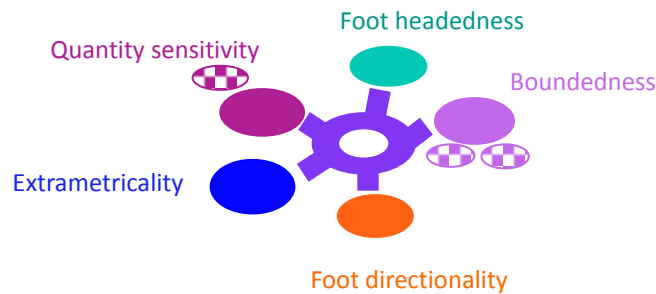


# Big picture



Establishing a methodology for **quantitatively evaluating competing linguistic knowledge representations**

- based on an argument from acquisition (how learnable are the language-specific grammars from realistic acquisition data?)



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



# Big picture

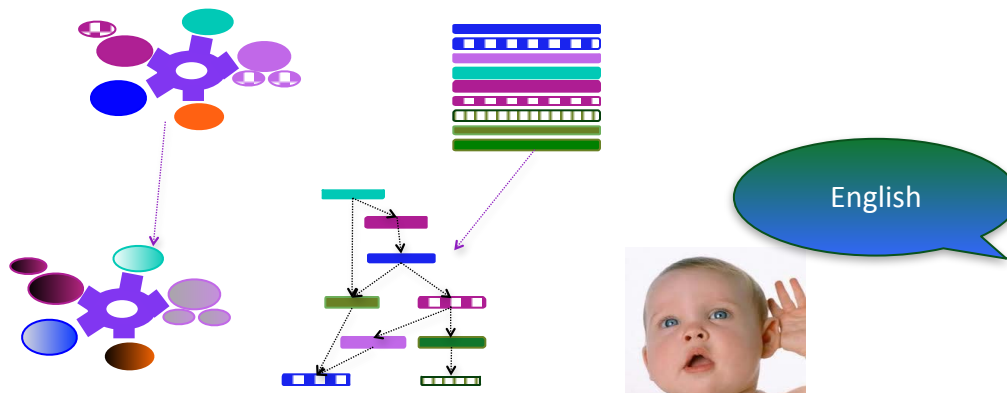


Establishing a methodology for **quantitatively evaluating** competing linguistic **knowledge representations**

- based on an argument from acquisition (how learnable are the language-specific grammars from realistic acquisition data?)

Testing against hard cases, like learning **English**

Idea: If a representation can handle the hard cases, it's really useful



# Big picture



Establishing a methodology for **quantitatively evaluating competing linguistic knowledge representations**

- based on an argument from acquisition (how learnable are the language-specific grammars from realistic acquisition data?)

Testing against hard cases, like learning **English**

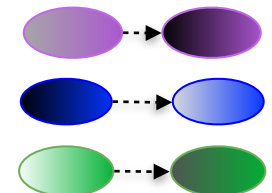
Idea: If a representation can handle the hard cases, it's really useful

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



Joanna Lee

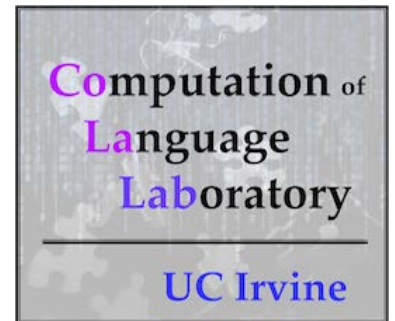
Pranav Anand  
Armin Mester

Adrian Brasoveanu  
Matt Wagers

Sandy Chung

Jim McCloskey

UC Santa Cruz Linguistics colloquium 2014  
Logic & Philosophy of Science 2013 colloquium, UC Irvine  
Institute for Mathematical Behavioral Sciences 2013, UC Irvine



Extra material

# Three knowledge representations

## Parametric systems

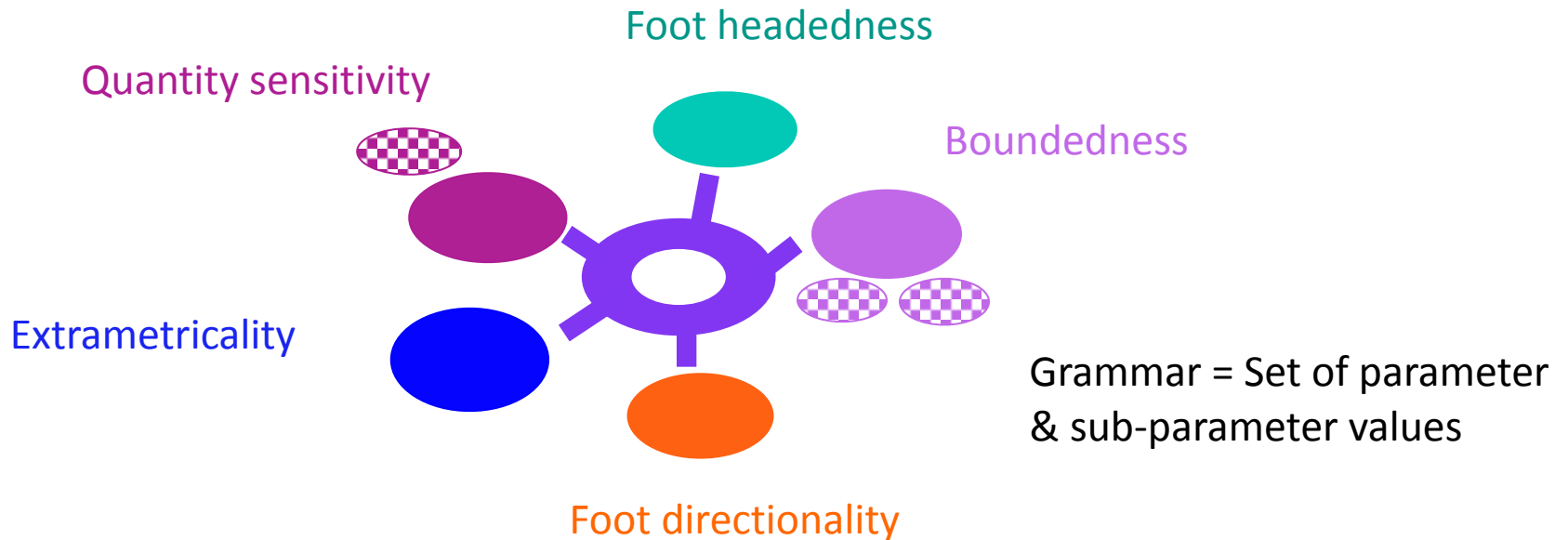
**HV**: Halle & Vergnaud 1987, Dresner 1999, Pearl 2011

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars

Correct grammar **builds**  
compatible contour

**O**ctopus



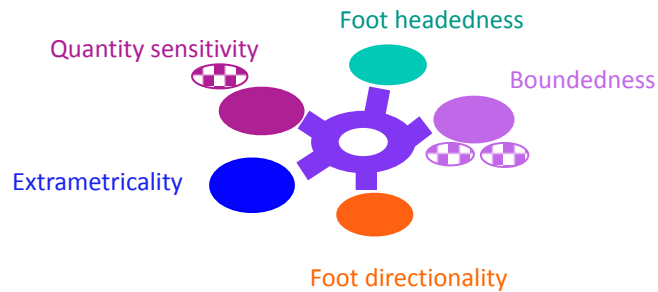
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Correct grammar builds compatible contour

OCTopus

oc to pus

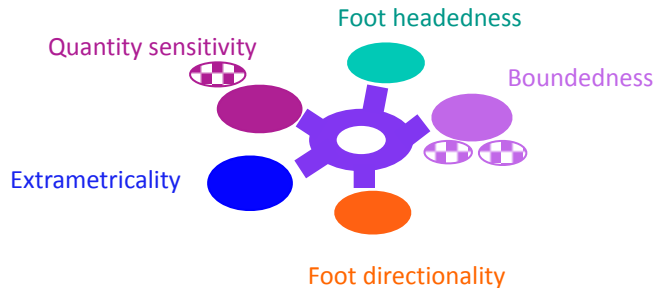
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



### Quantity sensitivity

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

Correct grammar **builds** compatible contour

OCtopus

H L H  
oc to pus



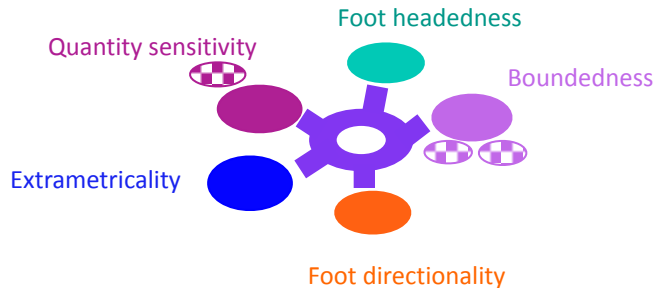
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



### Extrametricality

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

Correct grammar **builds** compatible contour

OCTopus

H L H  
oc to pus

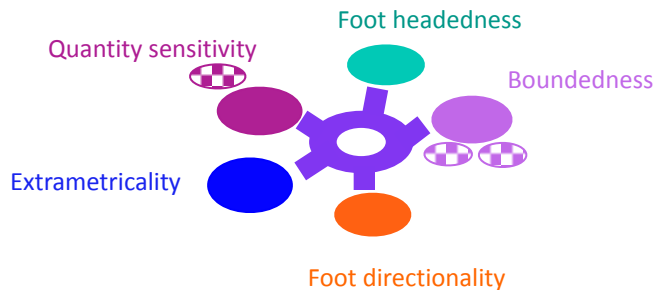
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



### Foot directionality

Are feet constructed from the left or from the right?

Correct grammar **builds**  
compatible contour

OCtopus

H L) (H  
oc to pus

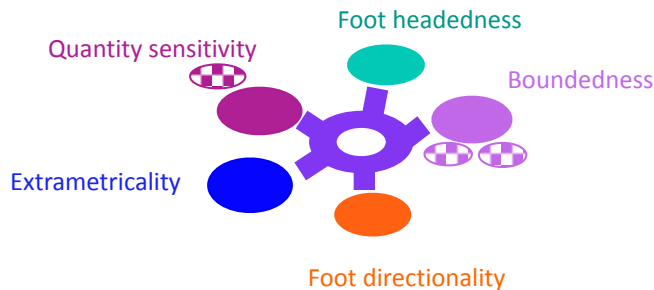
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



### Boundedness

How big are metrical feet?

Correct grammar **builds**  
compatible contour

OCTopus

(H L) ~~H~~  
oc to pus

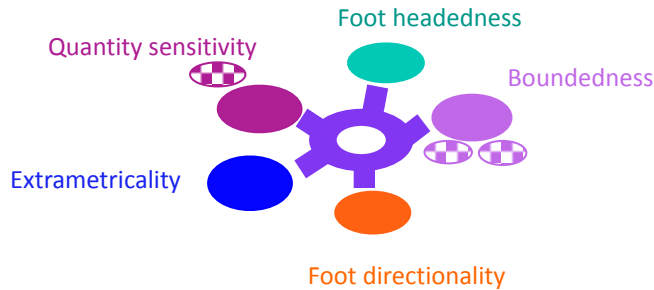
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



### Foot headedness

Which syllable in a foot is stressed?

Correct grammar **builds** compatible contour

OCTopus

(H L) H  
oc to pus

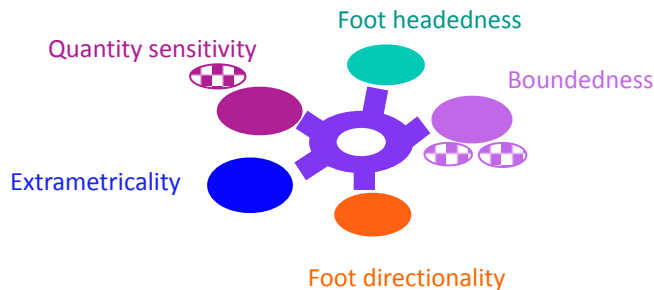
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



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

Correct grammar builds compatible contour

OCTopus

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

(H L) H  
OC to pus

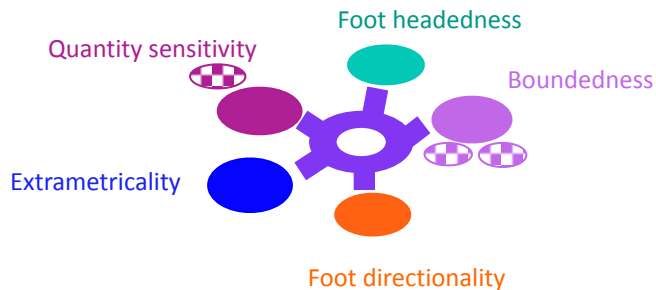
# Three knowledge representations

## Parametric systems

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

5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



Parameter values used:

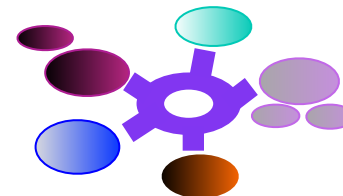
QS-VC-H, Em-Rt, FtDir-Rt, B-2-Syl, FtHd-Left

Correct grammar builds compatible contour

OCTopus

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

(H L) H  
OC to pus



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

# Three knowledge representations

## Parametric systems

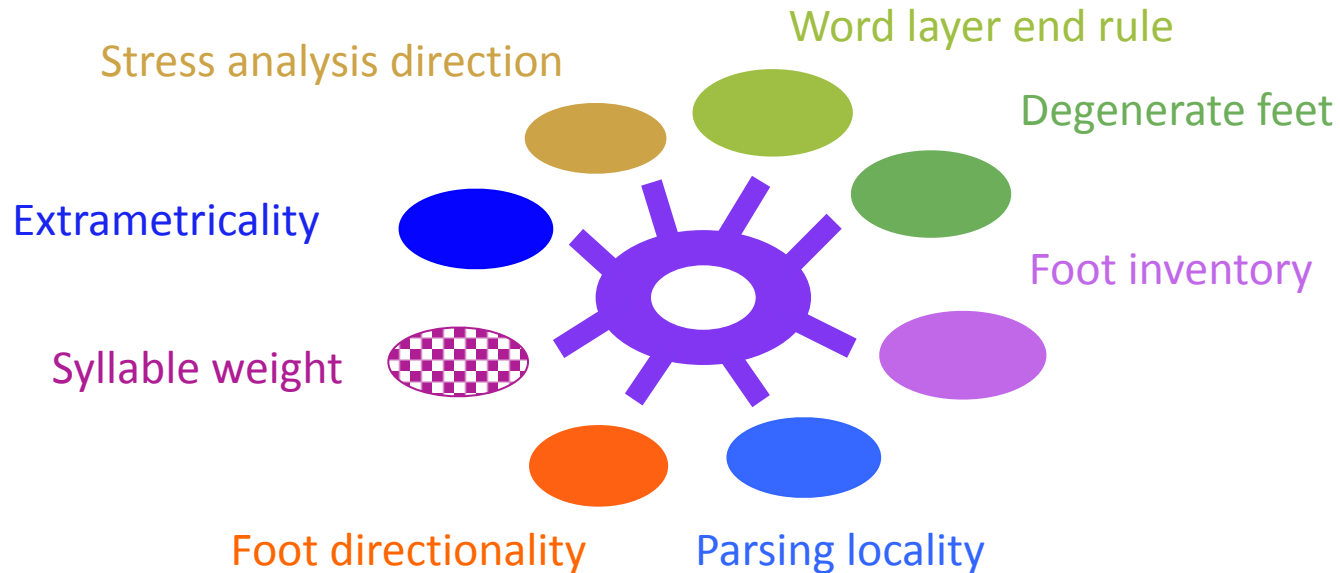
**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars

Correct grammar **builds**  
compatible contour

**O**ctopus



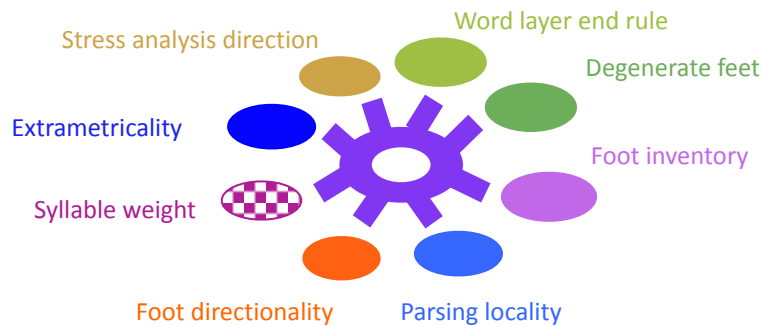
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Correct grammar **builds**  
compatible contour

**O**Ctopus

oc to pus



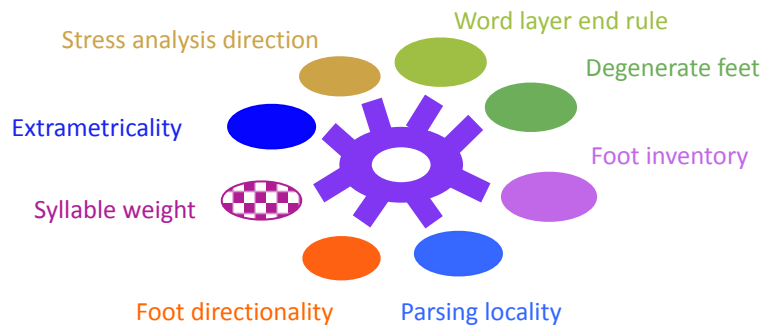
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Stress analysis direction

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

Correct grammar **builds** compatible contour

OCtopus

(...feet first...)

oc to pus

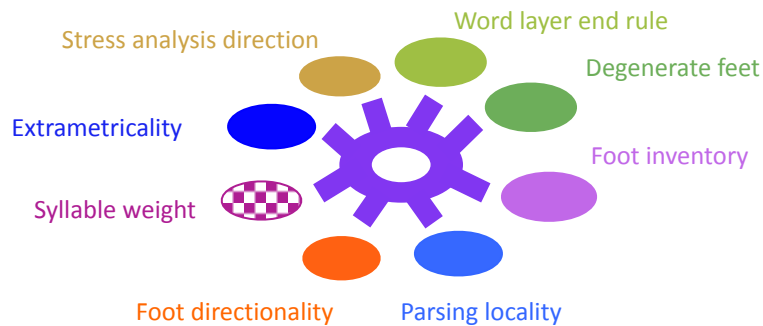
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Extrametricity

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

Correct grammar **builds** compatible contour

OCtopus

oc to pus

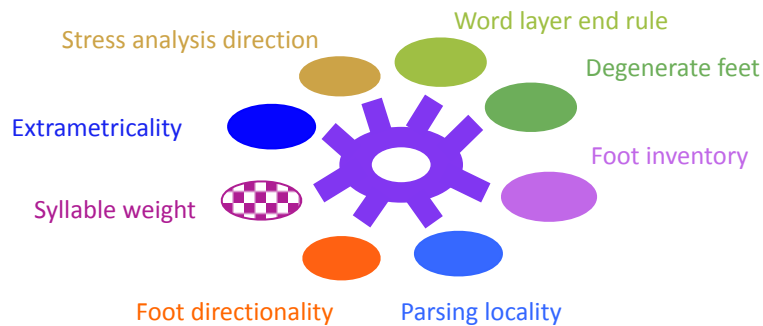
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Syllable weight

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

Correct grammar **builds** compatible contour

OCtopus

H L L  
oc to pus

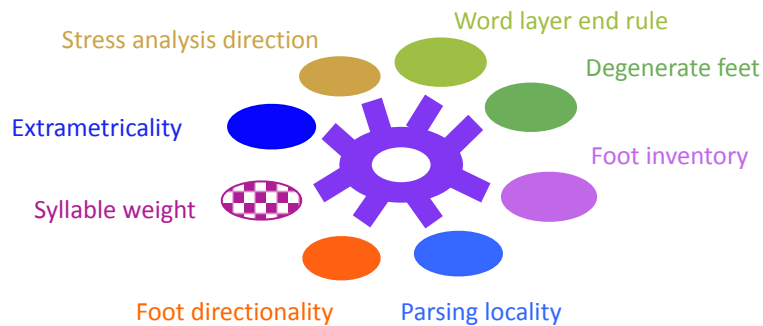
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Foot directionality

Are metrical feet constructed from the left or the right?

Correct grammar **builds** compatible contour

OCtopus

H L L )  
oc to pus

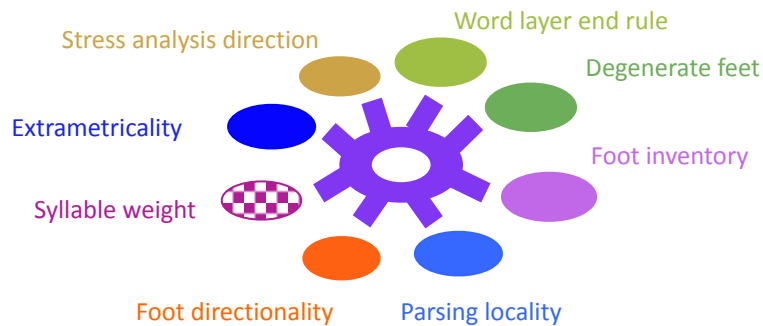
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Parsing locality

Is one Light syllable skipped between metrical feet?

Correct grammar **builds** compatible contour

OCtopus

?

H L L  
oc to pus

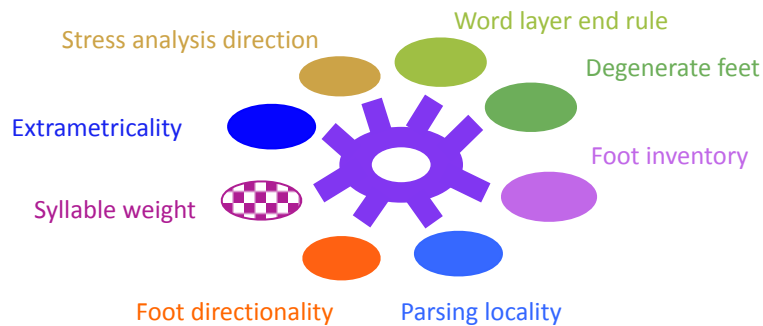
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

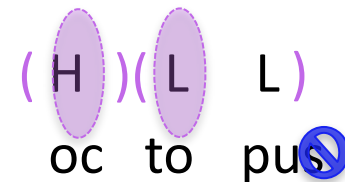
8 parameters

Hypothesis space: 768 grammars



Correct grammar **builds** compatible contour

OCtopus



## Foot inventory

How big are metrical feet?

Where does the stress fall within them?

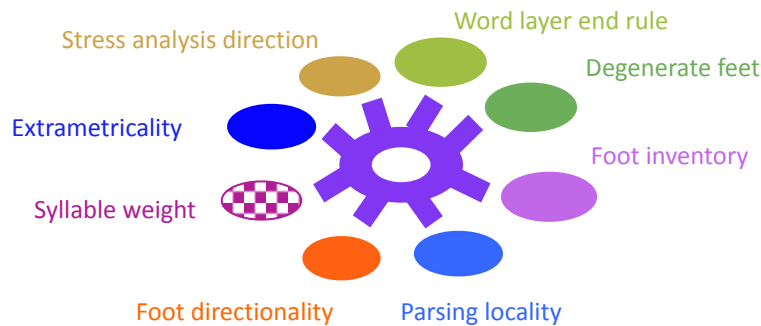
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Degenerate feet

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

Correct grammar **builds** compatible contour

OCtopus

( H )( L L )  
oc to pus

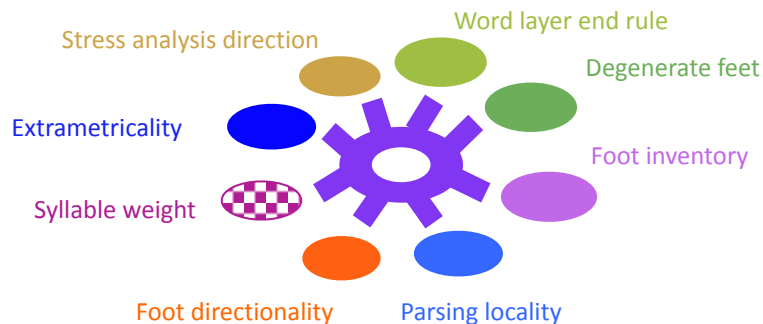
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



### Word layer end rule

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 L )  
oc to pus



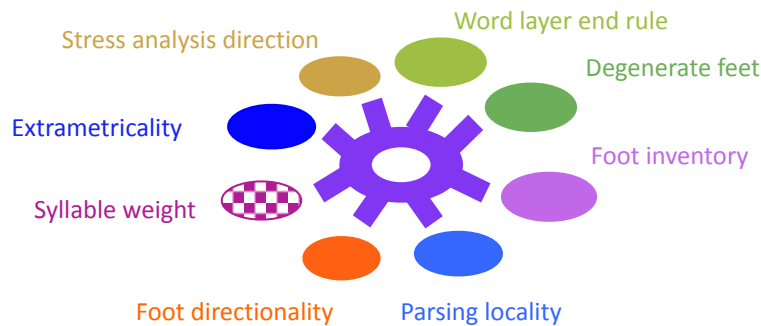
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Parameter values used:

Bottom-up, Extrametricity 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

Correct grammar builds compatible contour

OCtopus

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

( H )( L L )  
OC TÓ pus

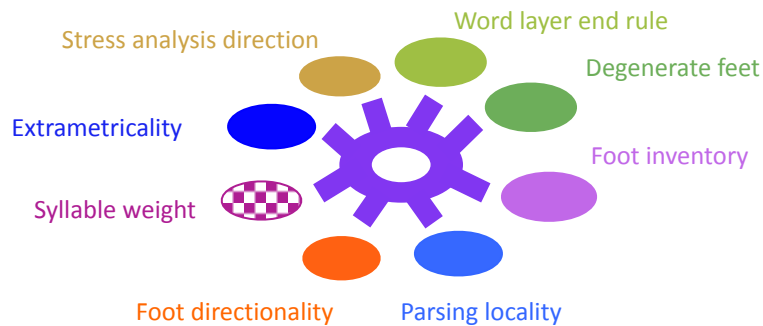
# Three knowledge representations

## Parametric systems

**Hayes:** Hayes 1995

8 parameters

Hypothesis space: 768 grammars



Parameter values used:

Bot, Em-RtCons, VC-H, FtDir-Rt,  
PL-Strong, MorTro, DF-Strong, WLER-Rt

Correct grammar builds compatible contour

OCTopus

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

( H )( L L )  
OC TÓ pus



...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

**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.

Best candidate for the correct grammar has a compatible contour

Octopus

(OC to) pus

oc (TO pus)

(oc TO) pus

	C1	C2	C3	C4
(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

Best candidate for the correct grammar has a compatible contour

OCTopus

Grammar = ranked ordering of all constraints



# Three knowledge representations

## Constraint-ranking systems

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

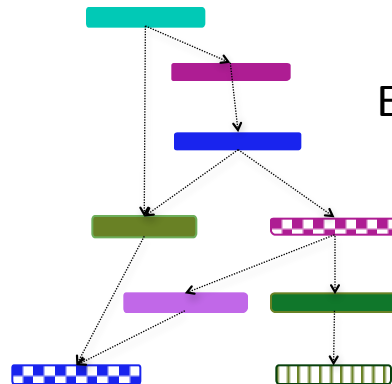
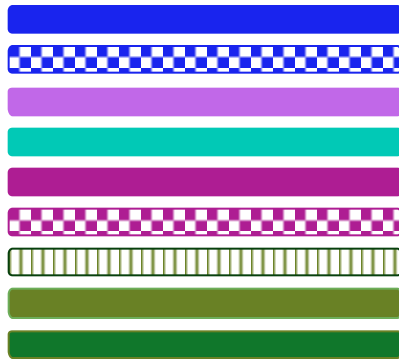
9 violable constraints

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

Best candidate for the correct grammar has a compatible contour

Octopus

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



English grammar

# Three knowledge representations

## Constraint-ranking systems

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

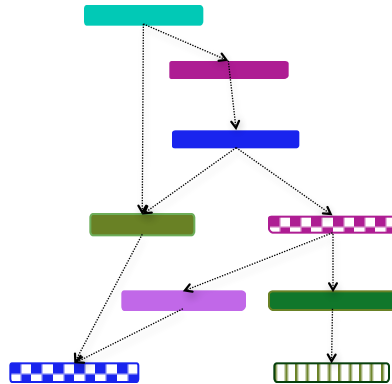
9 violable constraints

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

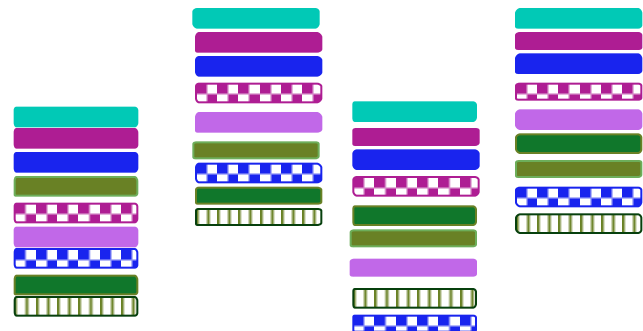
Best candidate for the correct grammar has a compatible contour

OCTopus

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



# 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

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

**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

## Parse- $\sigma$

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

**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

## 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

**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

## 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

**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

BAby

## Weight-to-Stress (VV)

Should all VV syllables be stressed?

(ba BY)

(BA) (BY)

(BA) by

(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


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

Best candidate for the correct grammar has a compatible contour

O**C**topus

## 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

**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

### 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

**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

### 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

**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- $\sigma$
-  Foot binarity
-  Trochaic
-  Trochaic
-  Weight-to-Stress
-  Align left, Align right
-  \*Sonorant nucleus

Best candidate for the correct grammar has a compatible contour

your**SELF**

### \*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

**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”:



## 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



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

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)

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



# 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



A sample grammar that is a version of the English “grammar”:



Next important: VV syllables are stressed.

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



A sample grammar that is a version of the English “grammar”:



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

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



A sample grammar that is a version of the English “grammar”:



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)

# 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

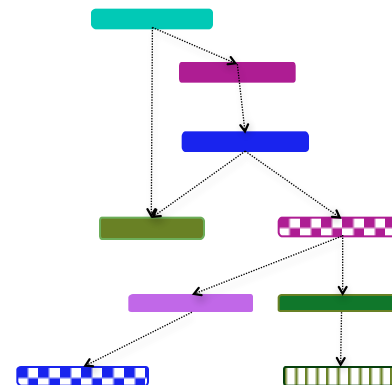
-  Nonfinality, Parse- $\sigma$
-  Foot binarity
-  Foot binarity
-  Trochaic
-  Weight-to-Stress
-  Weight-to-Stress
-  Align left, Align right
-  \*Sonorant nucleus
-  \*Sonorant nucleus

A sample grammar that is a version of the English “grammar”:



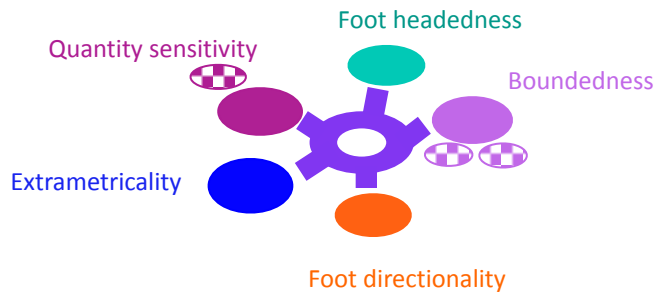
Best candidate for the correct grammar has a compatible contour

**O**Ctopus



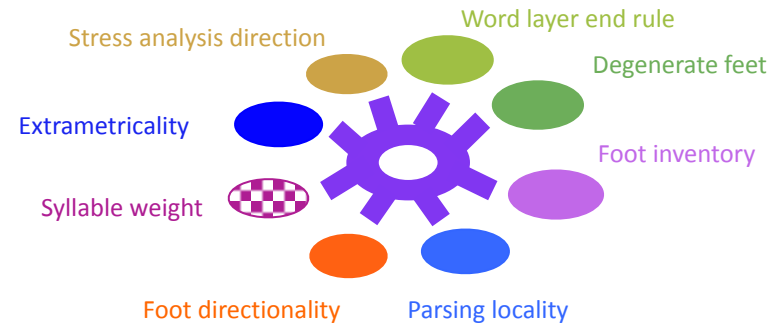
English “grammar”

# Knowledge representation comparison



**HV:** 5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



**Hayes:** 8 parameters

Hypothesis space: 768 grammars

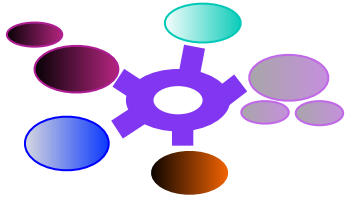


**OT:** 9 violable constraints

Hypothesis space: 362,880 grammars

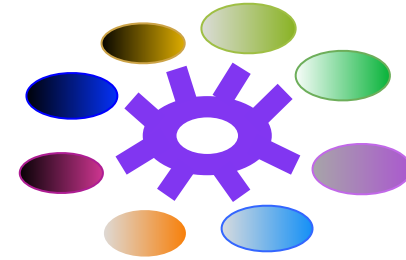


# English instantiations



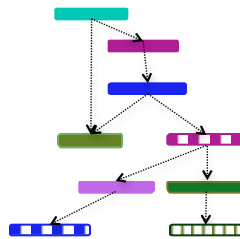
**HV:** 5 parameters & 4 sub-parameters

Hypothesis space: 156 grammars



**Hayes:** 8 parameters

Hypothesis space: 768 grammars

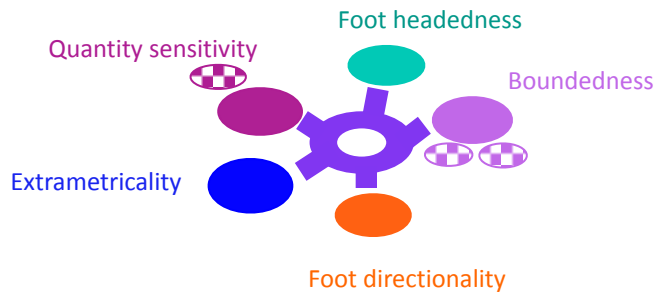


**OT:** 9 violable constraints

Hypothesis space: 362,880 grammars

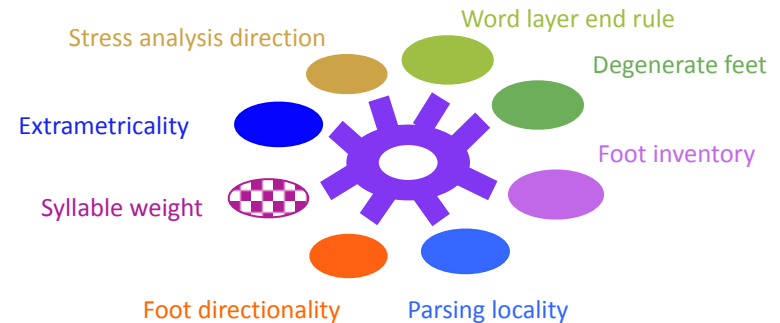
(English = 26 grammars)

# Knowledge representation comparison



**HV:** 5 parameters & 3 sub-parameters

Hypothesis space: 156 grammars



**Hayes:** 8 parameters

Hypothesis space: 768 grammars

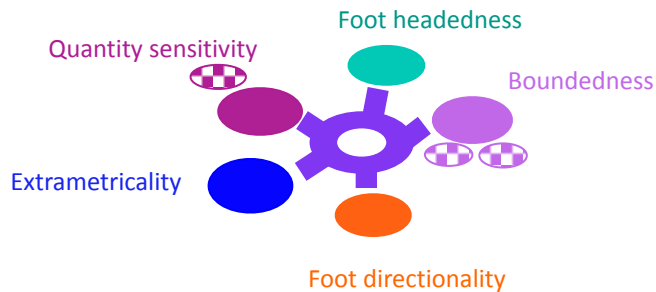
**Each representation assumes certain syllabic distinctions.**



**OT:** 9 violable constraints

Hypothesis space: 362,880 grammars

# Knowledge representation comparison

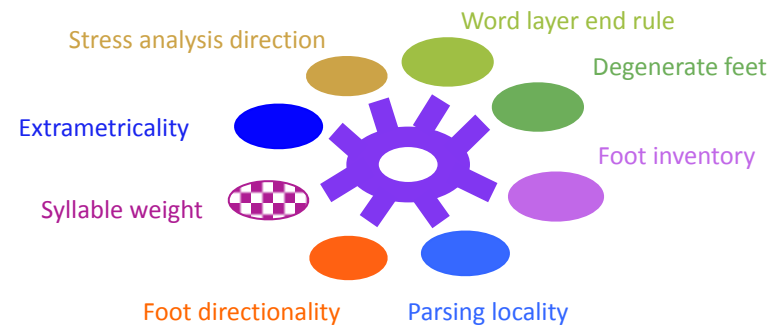


**HV:** 5 parameters & 3 sub-parameters

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)

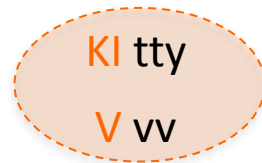
# Learning English metrical phonology: Non-trivial

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

Answer: **Syllabic learnability potential**

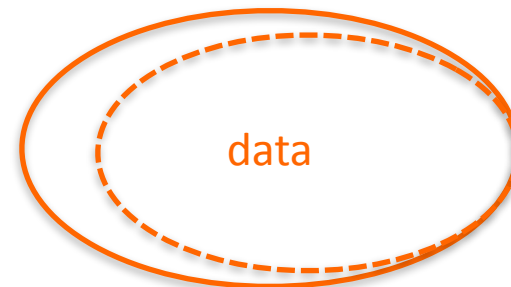
Account for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV



a WAY  
v VV

UH OH  
V VV



# 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

KI tty  
V vv

a WAY

v VV

UH OH

V VV

**Syllabic learnability potential by knowledge representation  
(proportion accounted for)**

HV: 0.711 types (0.766 tokens)

Hayes: 0.719 types (0.769 tokens)

OT: 0.795 types (0.829 tokens)

Pretty good potential coverage,  
even if it isn't perfect

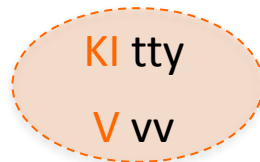
# Learning English metrical phonology

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

## Syllabic learnability potential

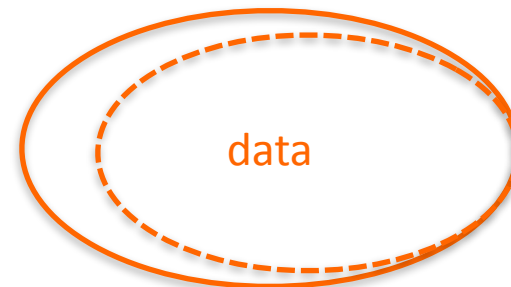
Accounting for the most frequent stress pattern for each syllabic word form.

Syllabic word form: V VV



a WAY  
v VV

UH OH  
V VV



# 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

## Syllabic learnability potential by knowledge representation (proportion accounted for)

HV: 0.708 types (0.766 tokens)

Hayes: 0.711 types (0.768 tokens)

OT: 0.820 types (0.852 tokens)

Still pretty good potential coverage...but it's also still not perfect.

# Learning English metrical phonology

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

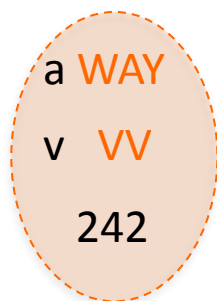
A grammar is constrained by the specific parameters or constraints defined by the KR. These parameters/constraints may not be compatible with the most frequent stress contour every time.

Syllabic word form: V VV

KI tty

V vv

18



Syllabic word form: VV VV

BA by

VV vv

198

o KAY

vv VV

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

KI tty

a WAY

V vv

v VV

18

242

Syllabic word form: VV VV

BA by

o KAY

VV vv

vv VV

198

4

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

KI tty  
V vv  
18

a WAY  
v VV  
242

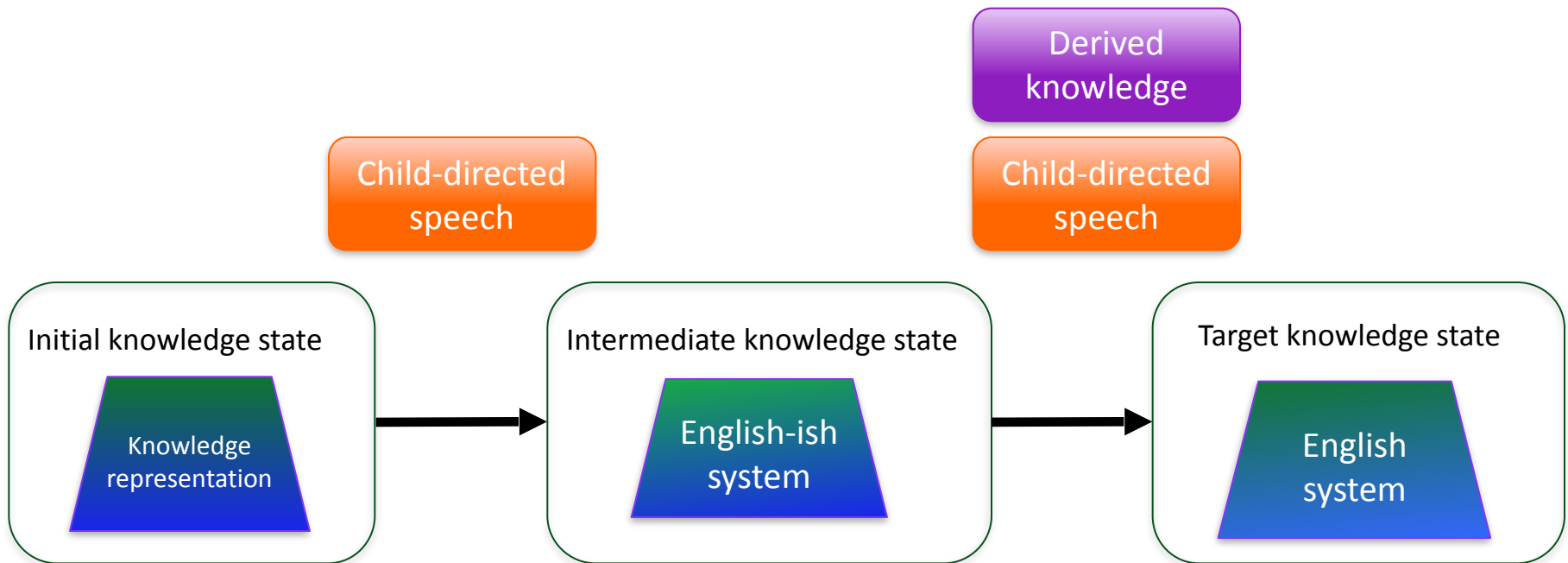
Syllabic word form: VV VV

BA by  
VV vv  
198

o KAY  
vv VV  
4

Grammar compatibility (preferring same contour): ...or these stress contours.

# 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.

**EAR**ly

**EAR**lier

**PRE**tty

**PRE**ttiest

sen**S**ation

sen**S**ational

sen**S**ationally

# 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

EARlier

PREtty

PREttiest

senSAtion

senSAtional

senSAtionally

# 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  $\approx 1.00$ ) 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  $\approx 1.00$ ) can account for

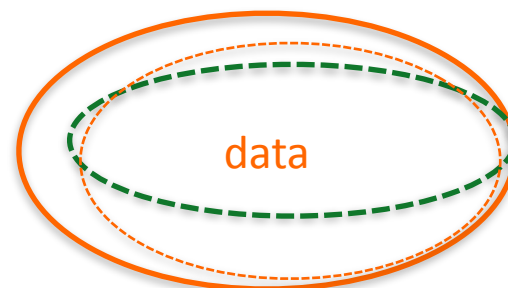
## **Raw compatibility of best grammar**

**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...**



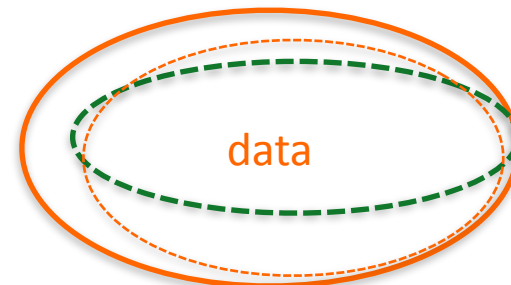
## Raw compatibility of best grammar

**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

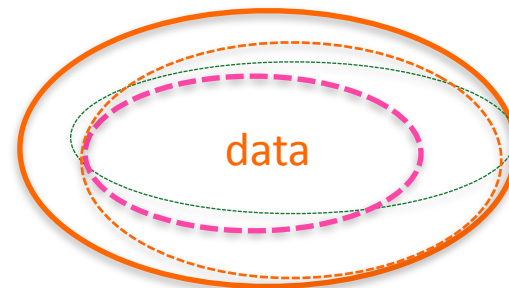
## Raw compatibility of English grammar

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)

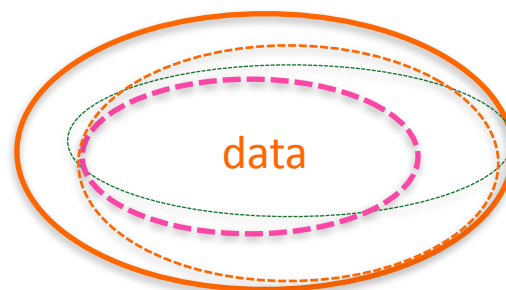
Still significantly less than the best grammar

## **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

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)

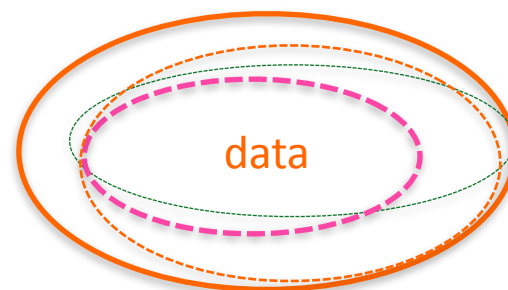
Still significantly less than the best grammar

## 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

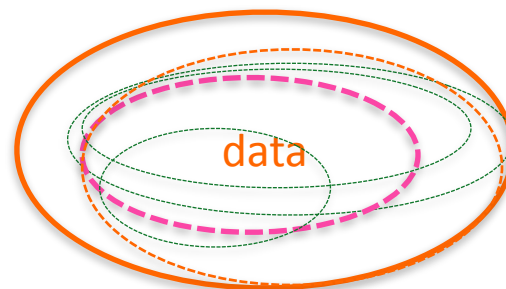
## Relative compatibility of English grammar

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:

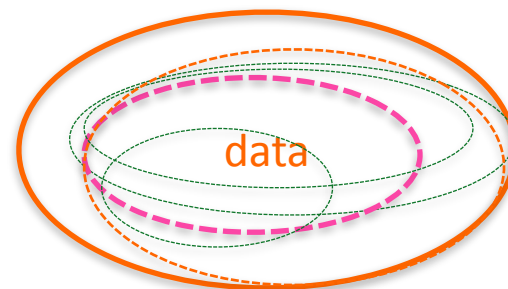
There remain many other grammars in the hypothesis space that are more compatible with the data, even though the learner knows inflectional morphology is stressless. Even if children aren't optimal learners, it would be easier to pick one of these other more compatible grammars.



## 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

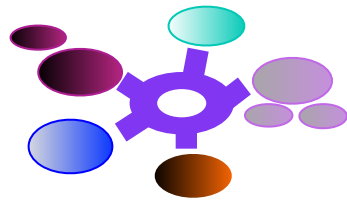
Better than many...but many are still better



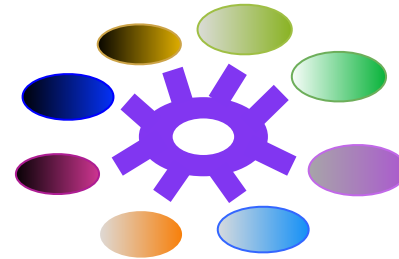
# Learning English metrical phonology

Continuing conclusion:

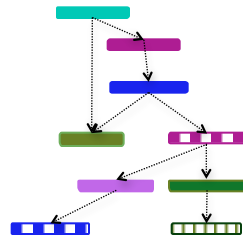
The same learnability issues persist for the English grammar in all three knowledge representations, even when the learner has some knowledge of the interactions between morphology and metrical phonology.



Parametric: HV



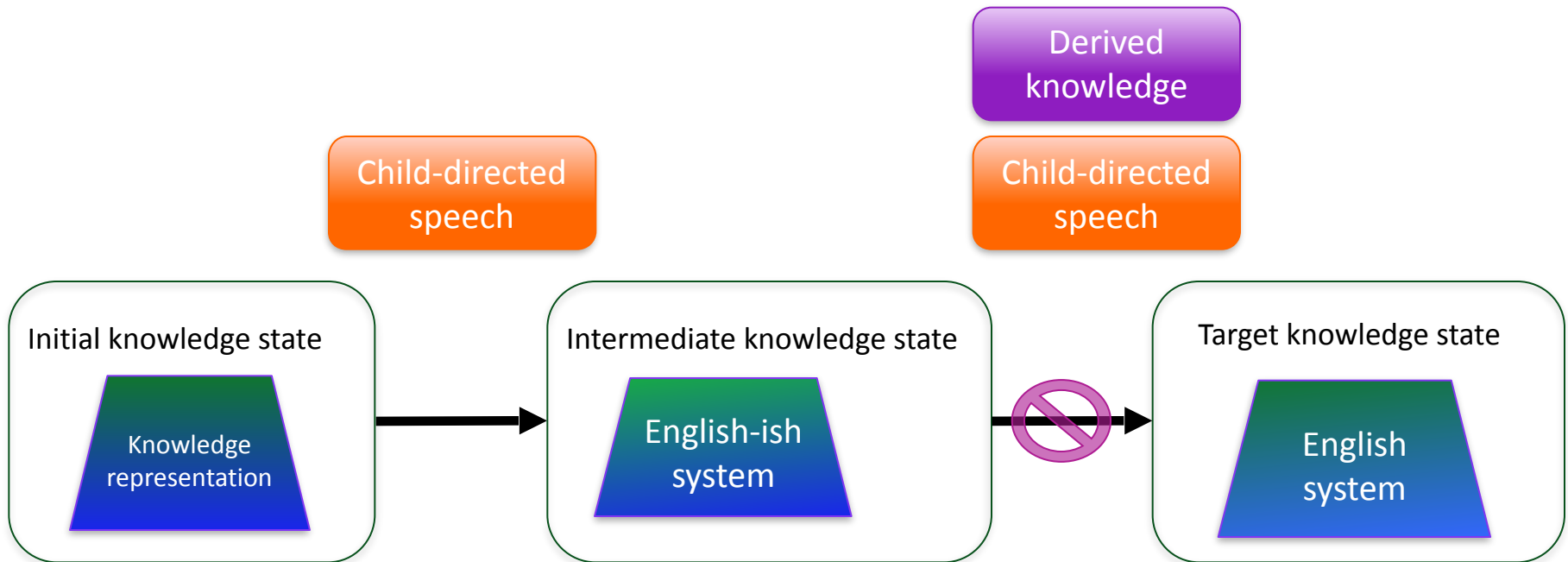
Parametric: Hayes



Constraint-based: OT



# 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 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

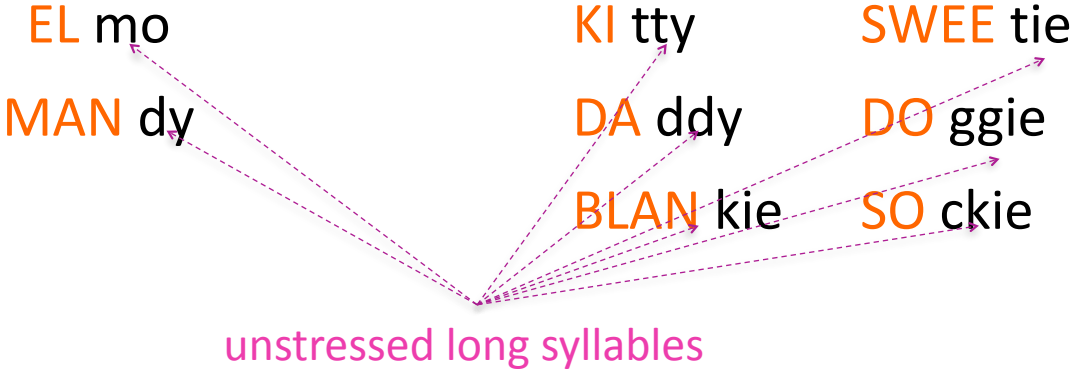
# The learnability problem: One option that didn't work

One problem:

All English grammars generally want **long syllables (VV 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**

**diminutives**



# The learnability problem: One option that didn't work

...yet

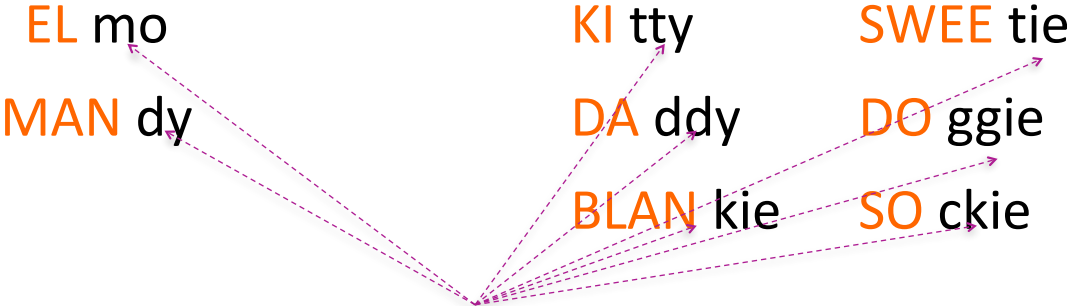
One solution: Additional knowledge

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



**proper names**

**diminutives**



unstressed long syllables

# The learnability problem: One option that didn't work

...yet

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**

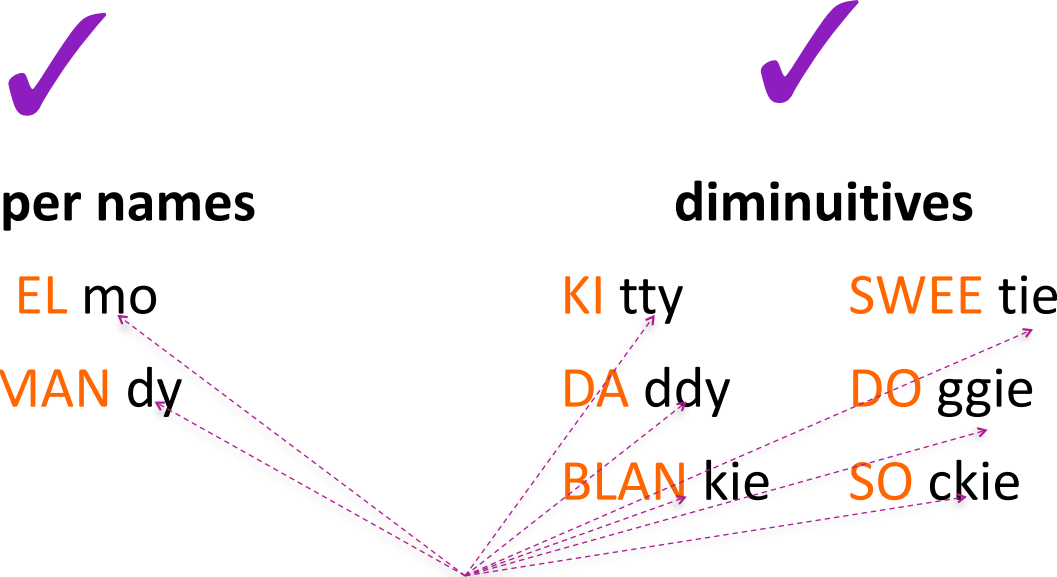
EL mo  
MAN dy



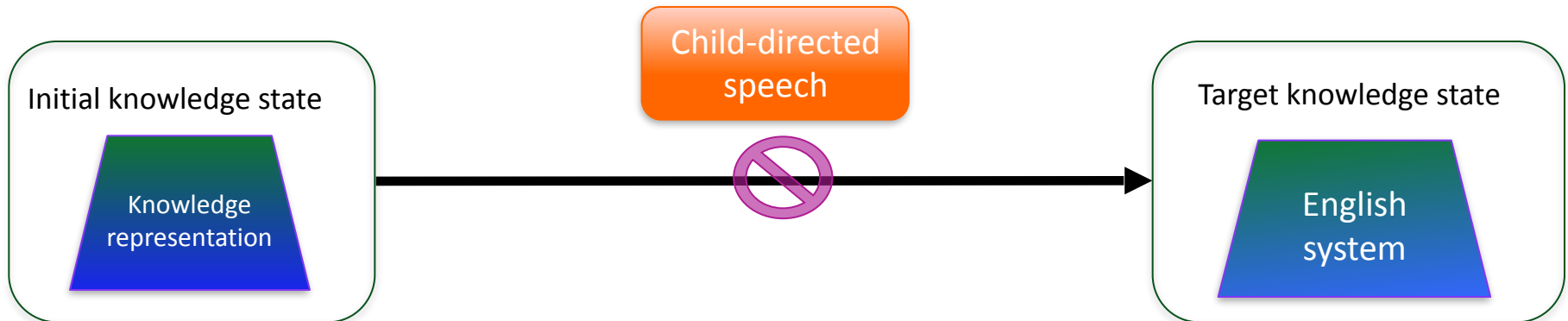
**diminutives**

KI tty      SWEE tie  
DA ddy      DO ggie  
BLAN kie      SO ckie

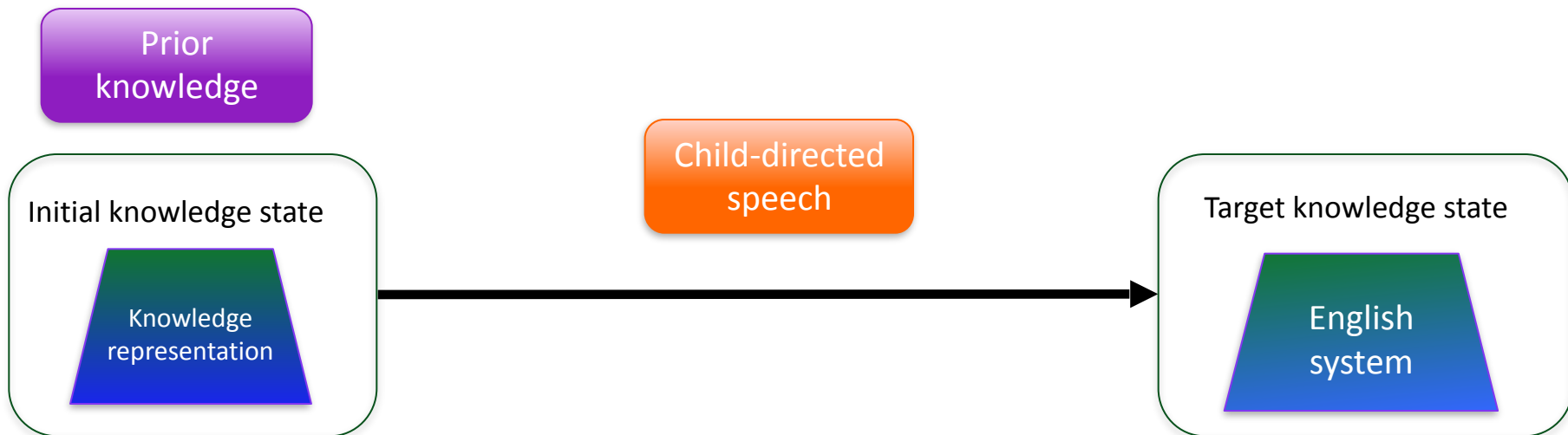
unstressed long syllables



# The learnability problem: Another option



# The learnability problem: Another option



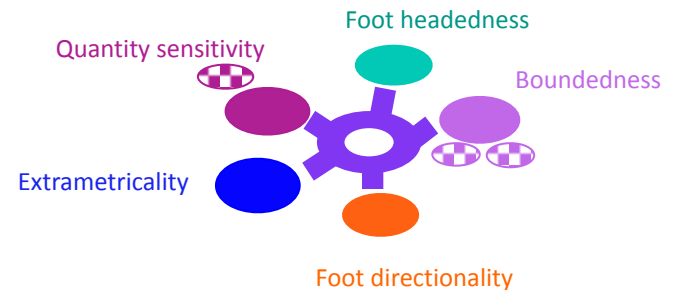
The initial state of the learner includes **prior knowledge** that helps the learner learn in a more sophisticated way.

# What kind of prior knowledge?

Helpful prior knowledge: **Learning biases**

Pearl 2008: The **parametric HV** English grammar can be learned from child-directed speech 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).





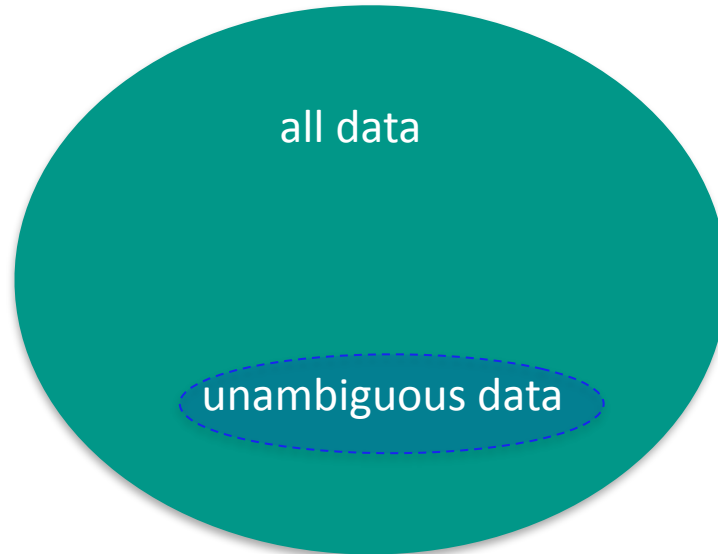
# Unambiguous data filter

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



# Unambiguous data filter

**Updated** working assumption: The learner will try to learn a grammar 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**.

# Unambiguous data filter

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

**Why** would this occur?

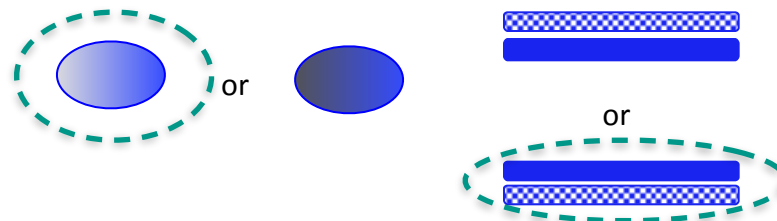
Perhaps the learner **prioritizes data that are viewed as highly informative**. The goal then becomes to learn a system that can account for all these data.

# Unambiguous data filter

**Updated** working assumption: The learner will try to learn a grammar 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 **productive** (Legate & Yang 2012).



COO kie

DA ddy

FU nny



a WAY

be LOW

to DAY



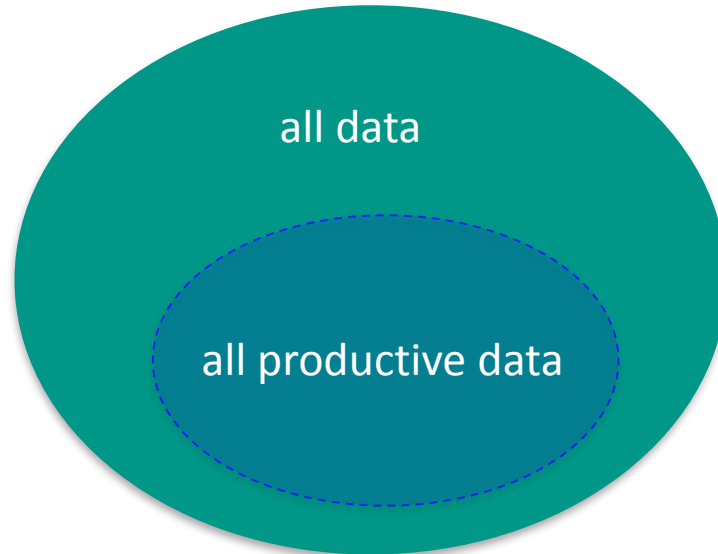
# Productive data filter

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



# Productive data filter

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



# Productive data filter

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

**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.

**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.

**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.

**How** would this occur?

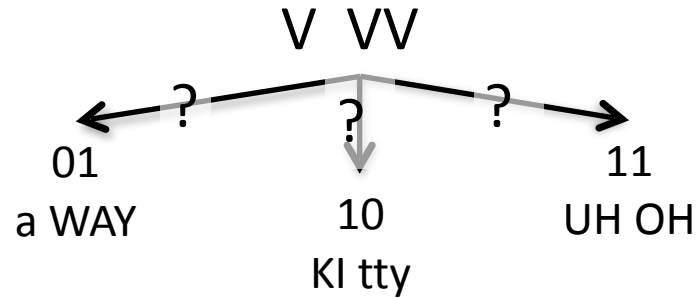
The number of exceptions a rule can tolerate for a set of  $N$  items is

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

(Yang 2005, Legate & Yang 2012)

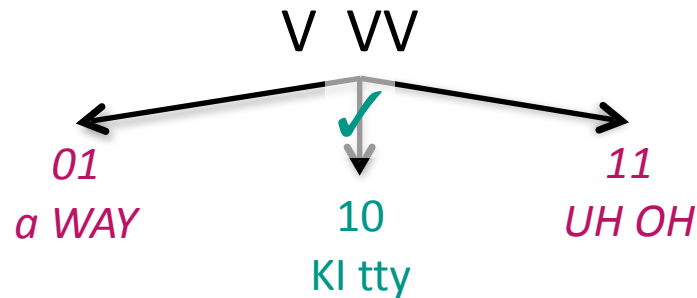
# The Tolerance Principle in action

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

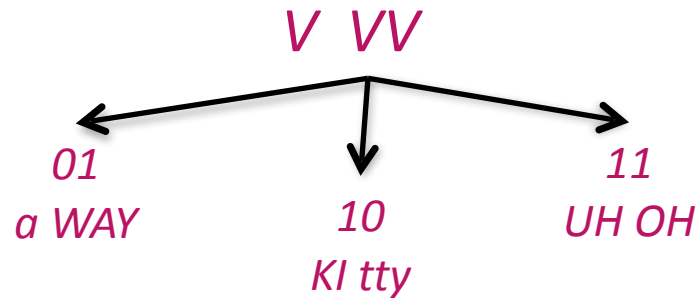
For every syllable word form with multiple stress contours, the learner could assess **whether any of those contours is the dominant one** (the “rule” for that syllable word form), using the Tolerance Principle.



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

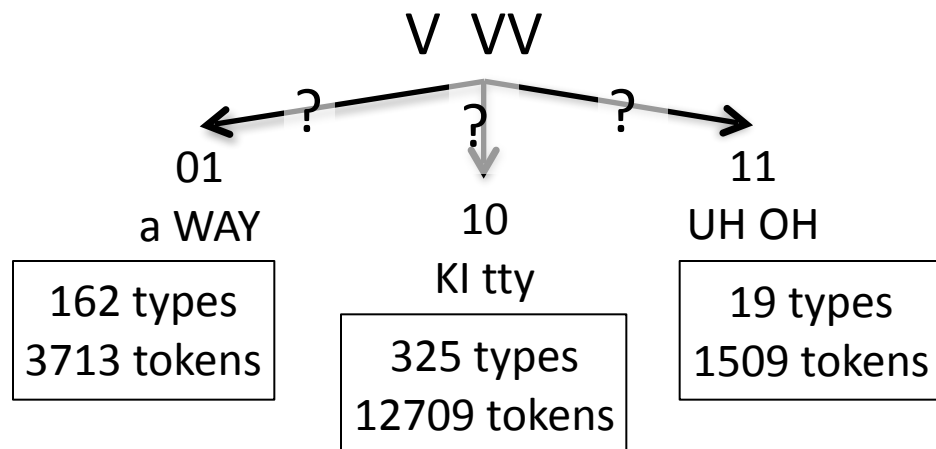
Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms

3 syllable distinctions

Hayes: 149 syllable word forms

4 syllable distinctions





# Productive data filter in action

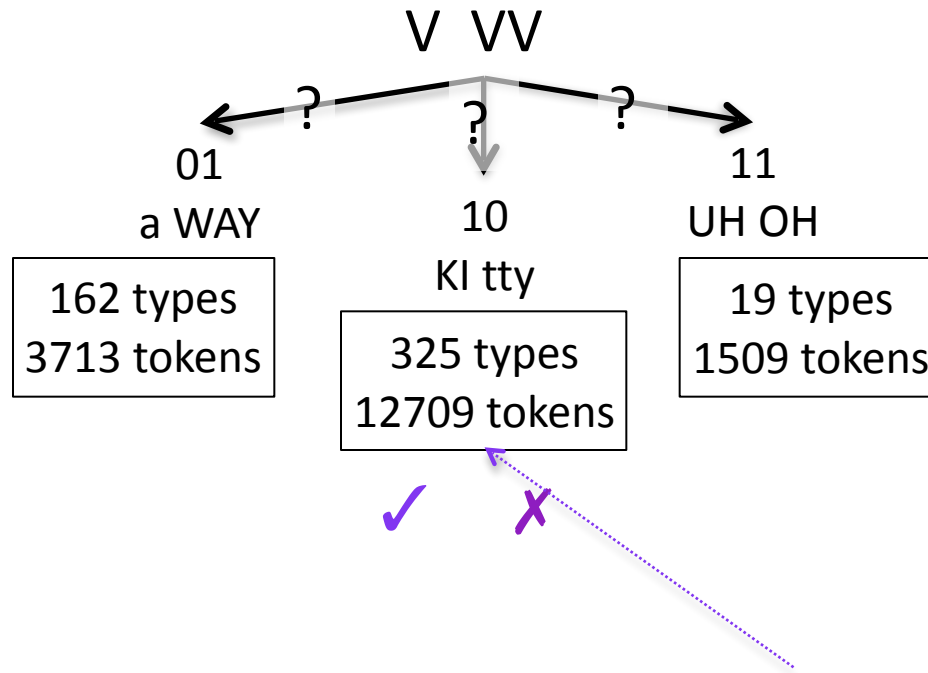
Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms

Hayes: 149 syllable word forms

3 syllable distinctions

4 syllable distinctions



These items are good for the HV English grammar.

# Productive data filter in action

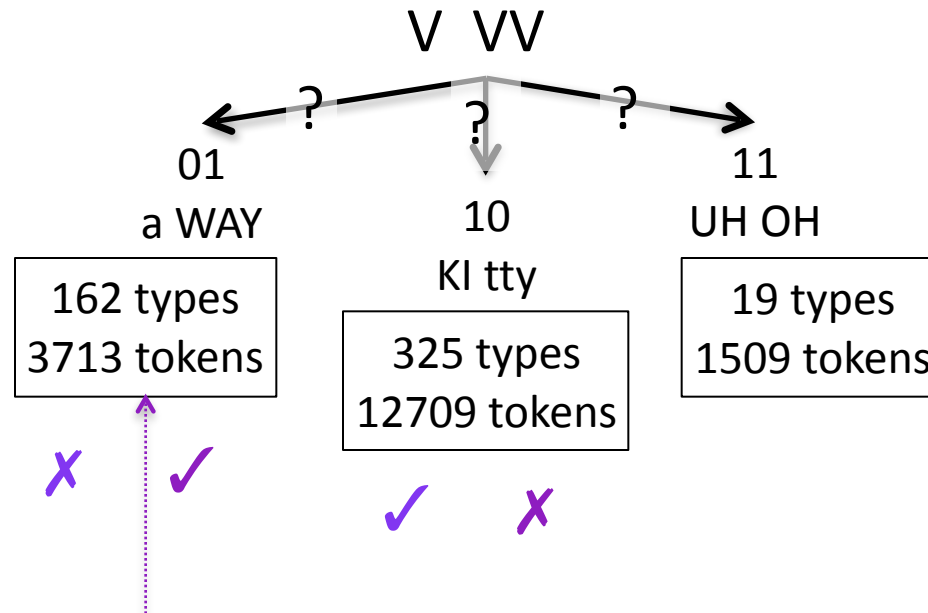
Parametric: **HV & Hayes**, with inflectional knowledge

HV: 123 syllable word forms

Hayes: 149 syllable word forms

3 syllable distinctions

4 syllable distinctions



These items are good for the Hayes English grammar.

# Productive data filter in action

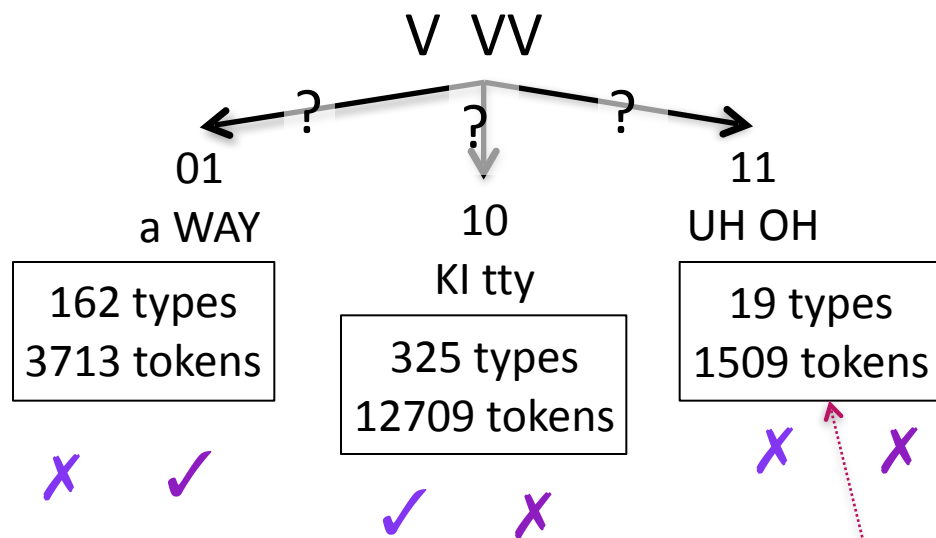
Parametric: **HV & Hayes**, with inflectional knowledge

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

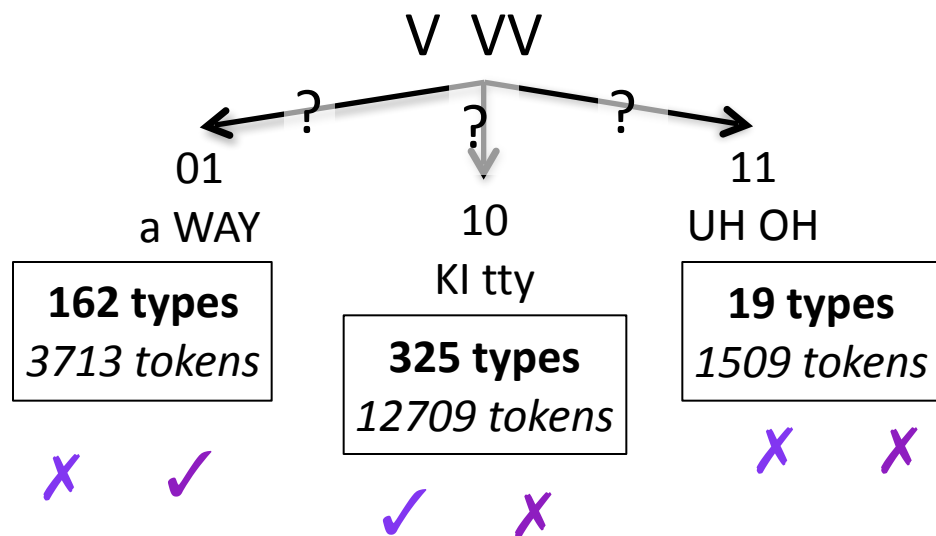
Parametric: **HV & Hayes**, with inflectional knowledge

HV: **123** syllable word forms

Hayes: **149** syllable word forms

3 syllable distinctions

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

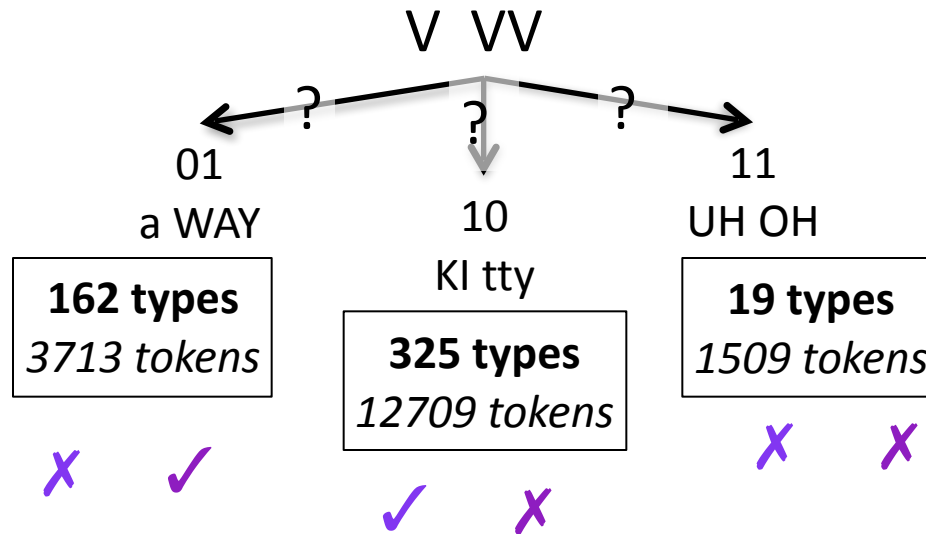
Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms

Hayes: 149 syllable word forms

3 syllable distinctions

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

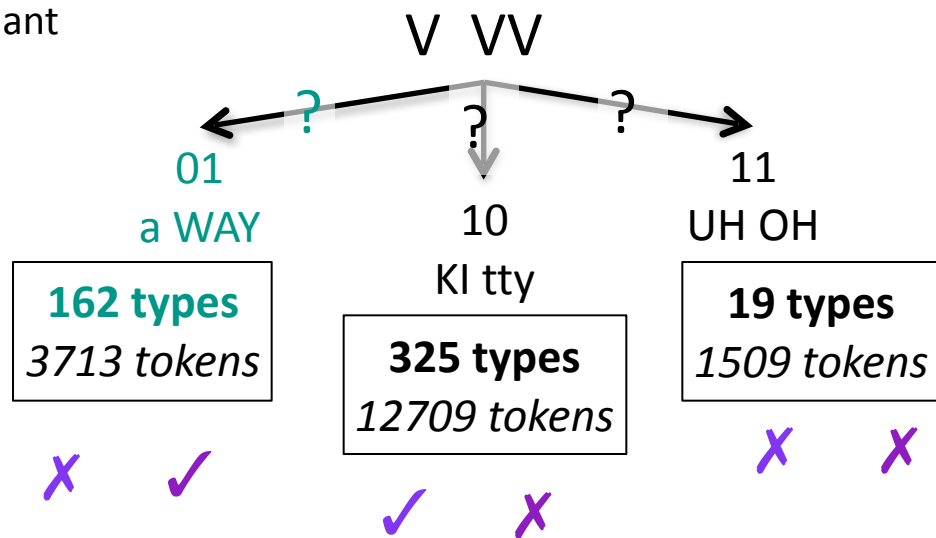
HV: 123 syllable word forms

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

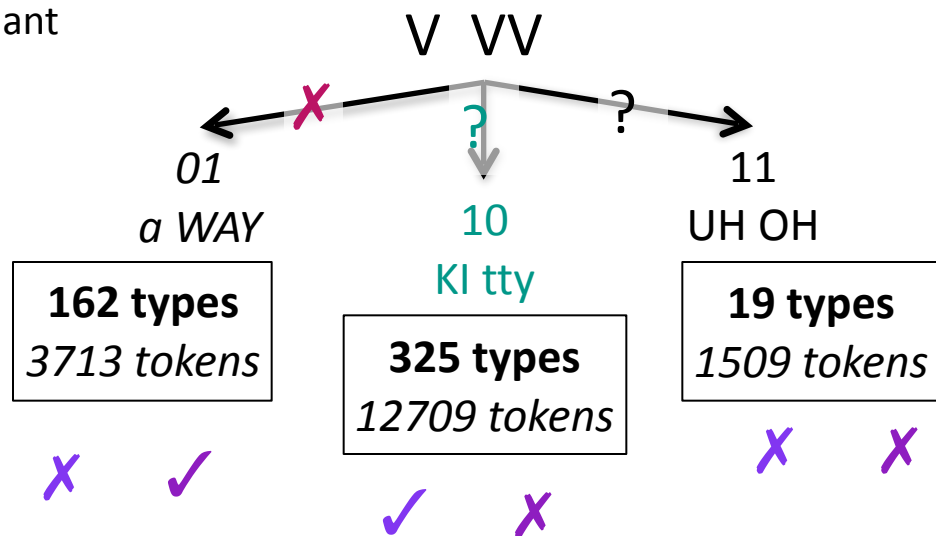
Hayes: 149 syllable word forms

3 syllable distinctions

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

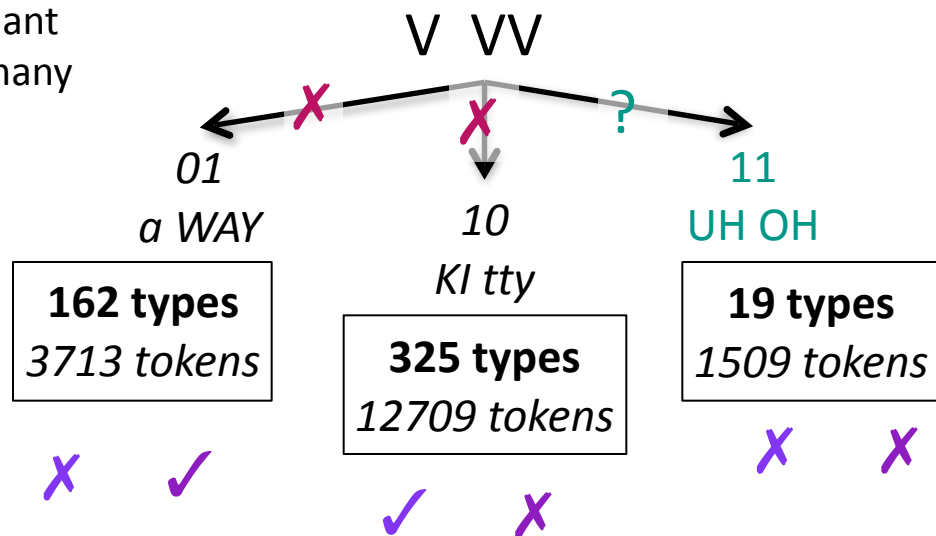
Hayes: **149** syllable word forms

3 syllable distinctions

4 syllable distinctions

If this is the dominant pattern, way too many exceptions:

$162 + 325 > 81$



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



# Productive data filter in action

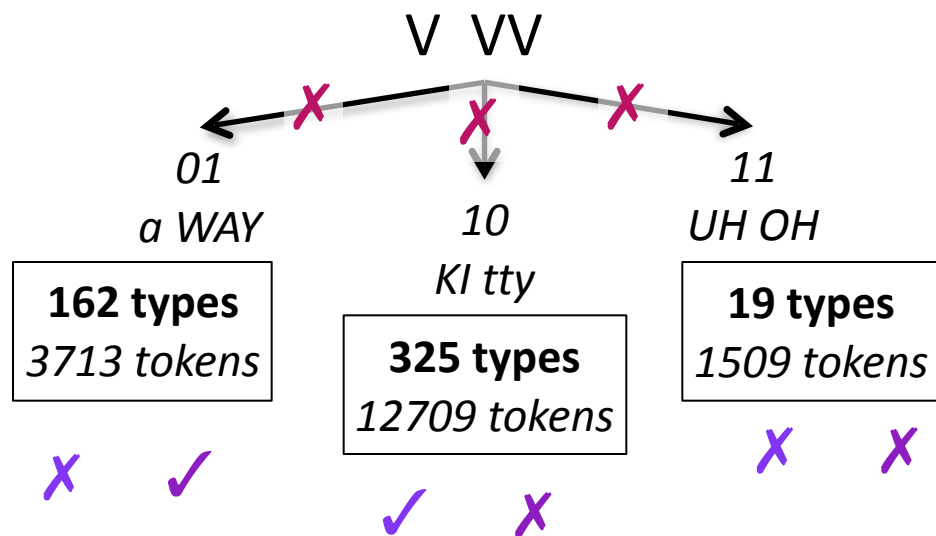
Parametric: **HV & Hayes**, with inflectional knowledge

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

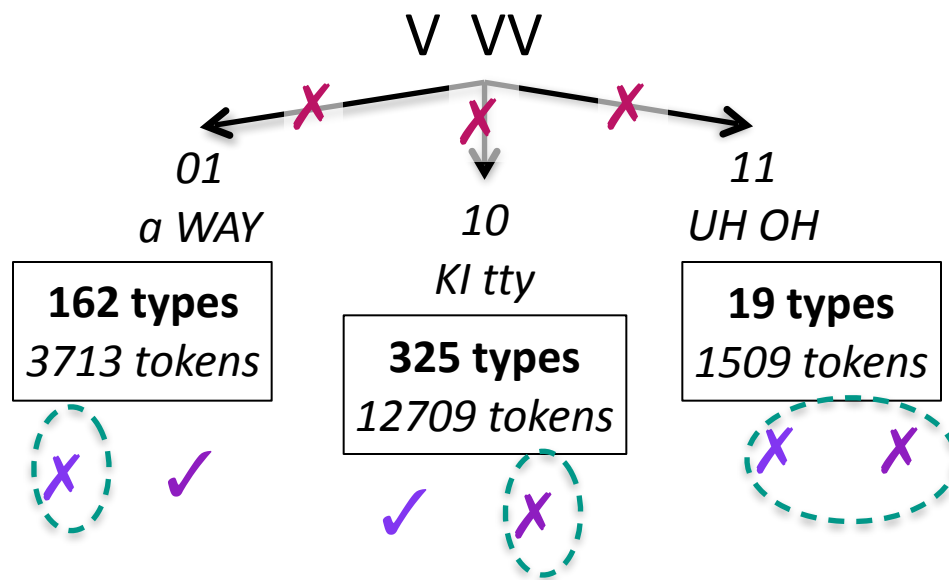
Parametric: HV & Hayes, with inflectional knowledge

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

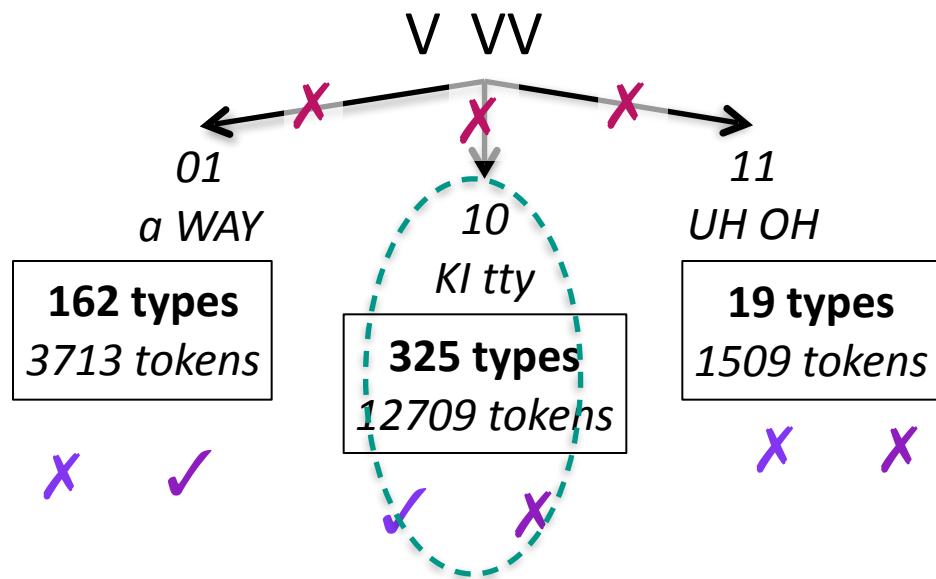
Parametric: HV & Hayes, with inflectional knowledge

HV: 123 syllable word forms

Hayes: 149 syllable word forms

3 syllable distinctions

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

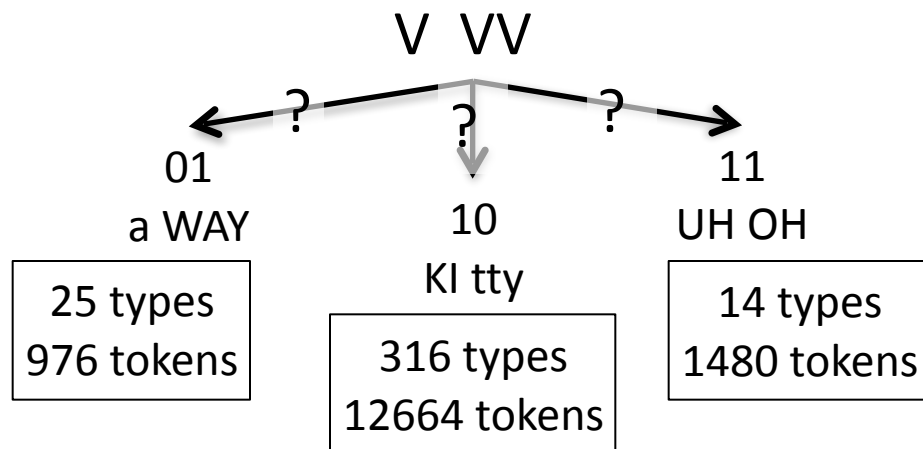
8 syllable distinctions

# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions

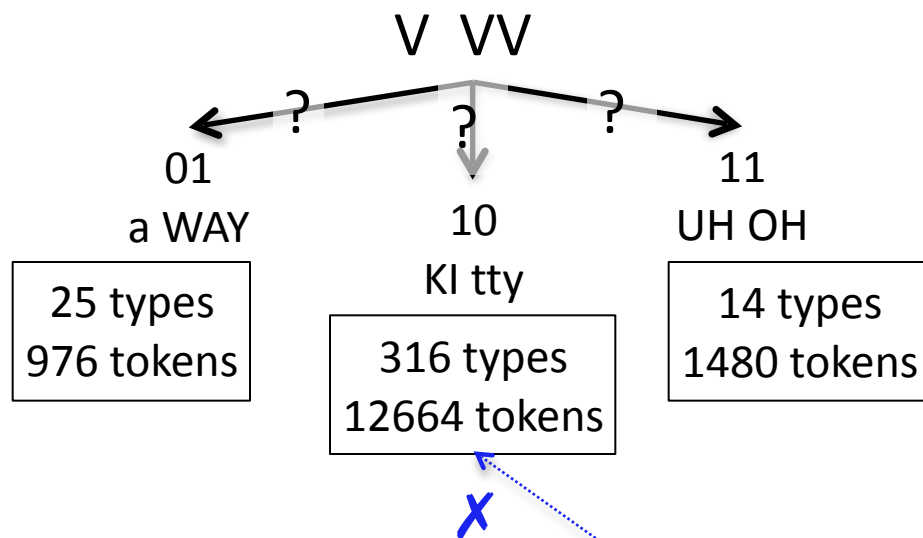


# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions



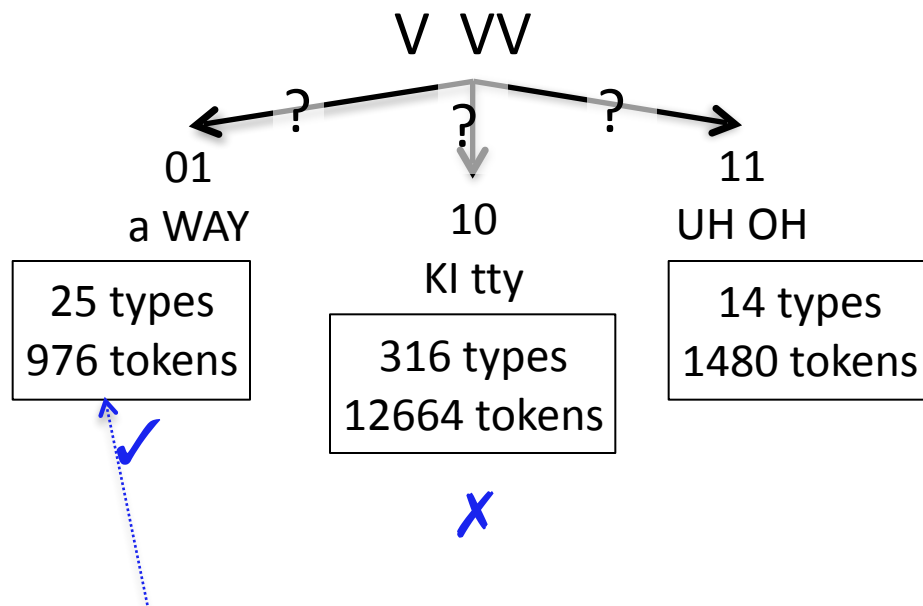
These items are bad for all English grammars.

# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions



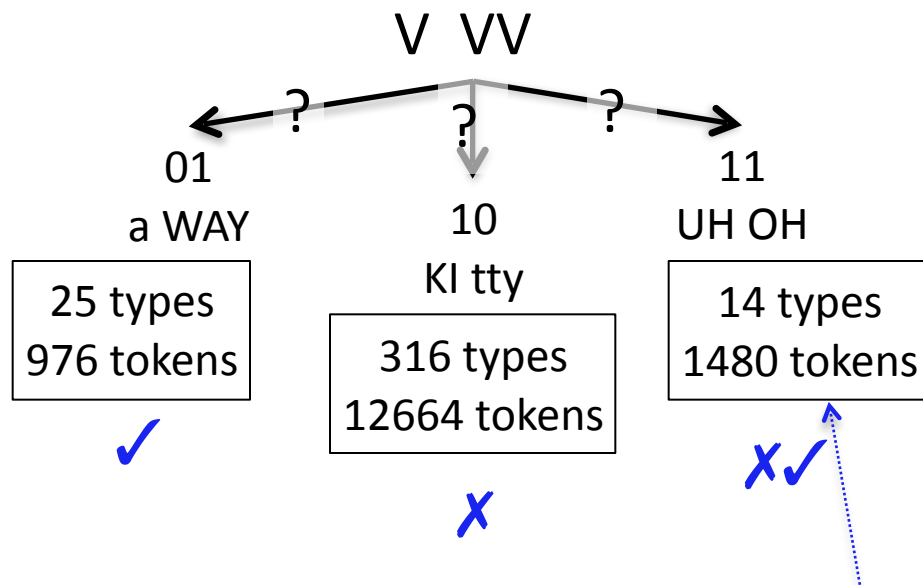
These items are good for most English grammars (21/26).

# 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).

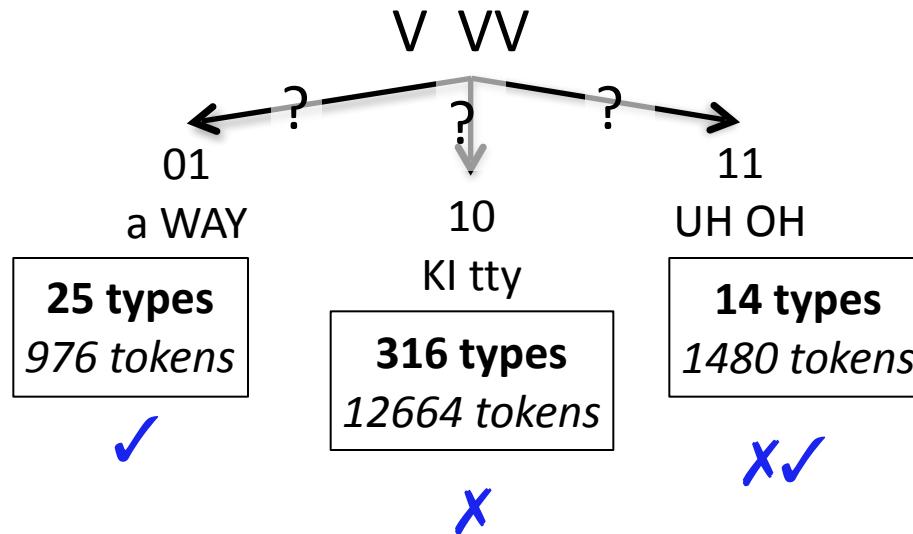


# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions



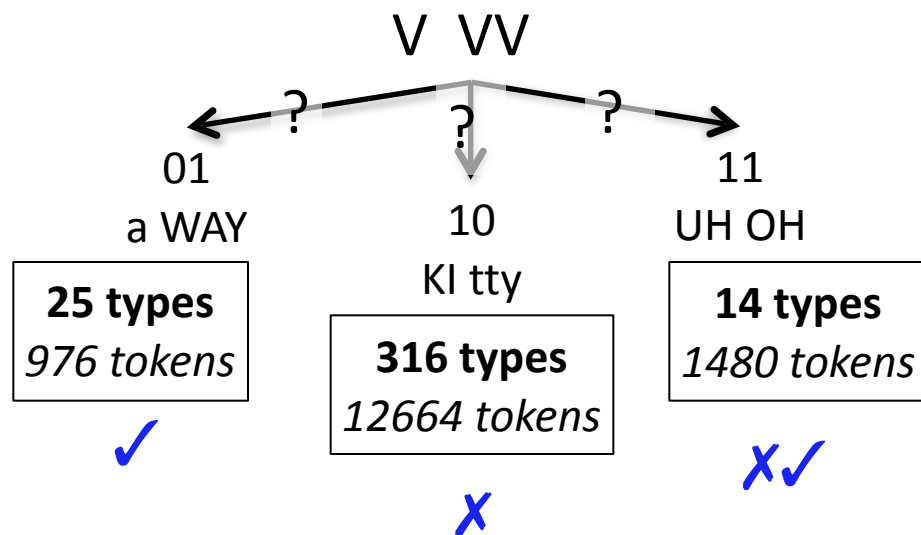
How many items should the stress “rule” apply to?  $N = 25 + 316 + 14 = 355$

# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions



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

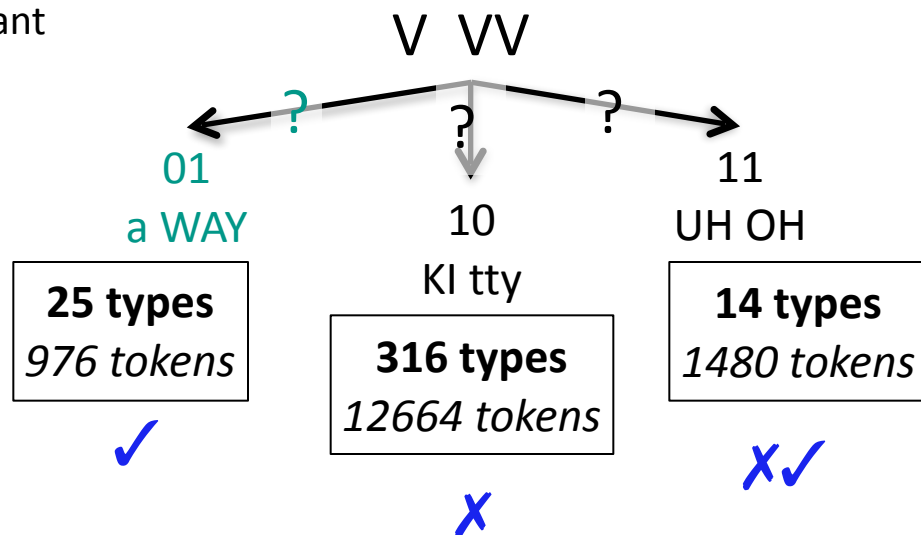
# Productive data filter in action

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

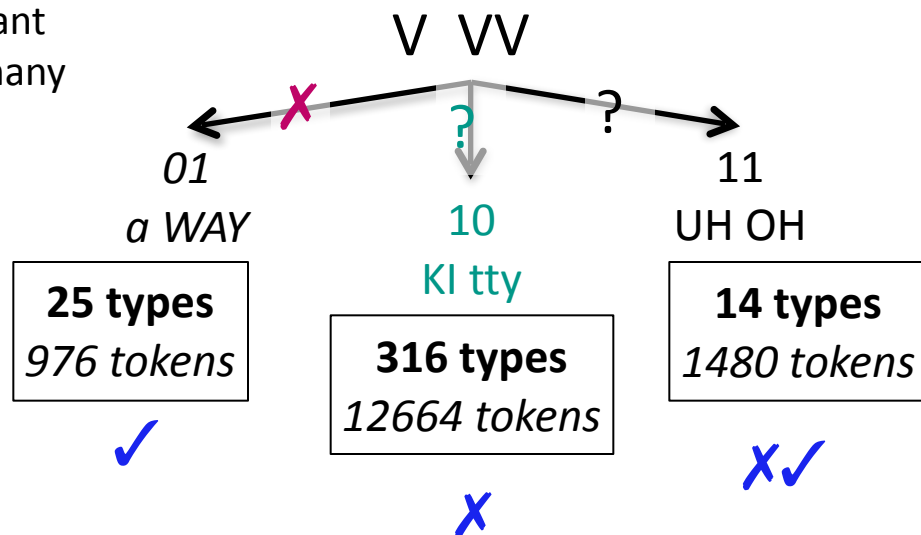
Constraint-based: **OT**, with inflectional knowledge

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?  $355 / \ln(355) = 60$

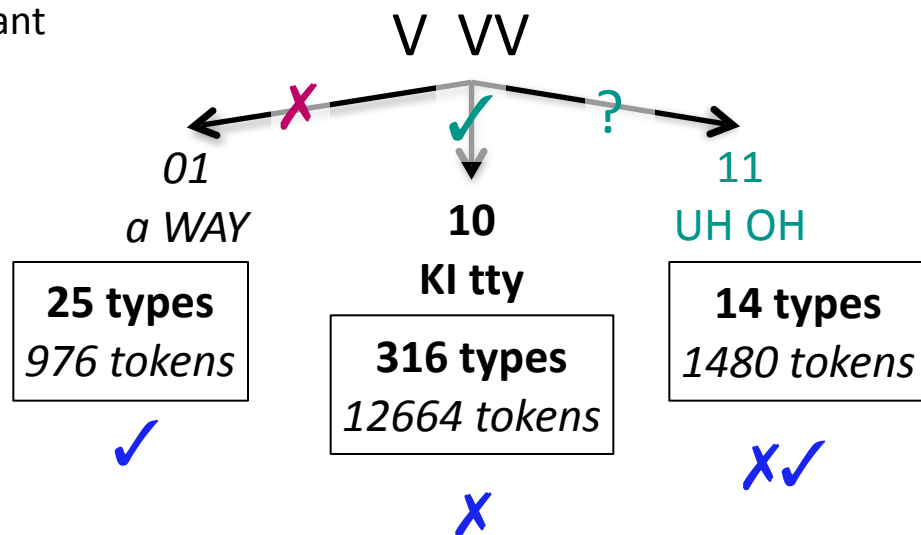
# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions

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



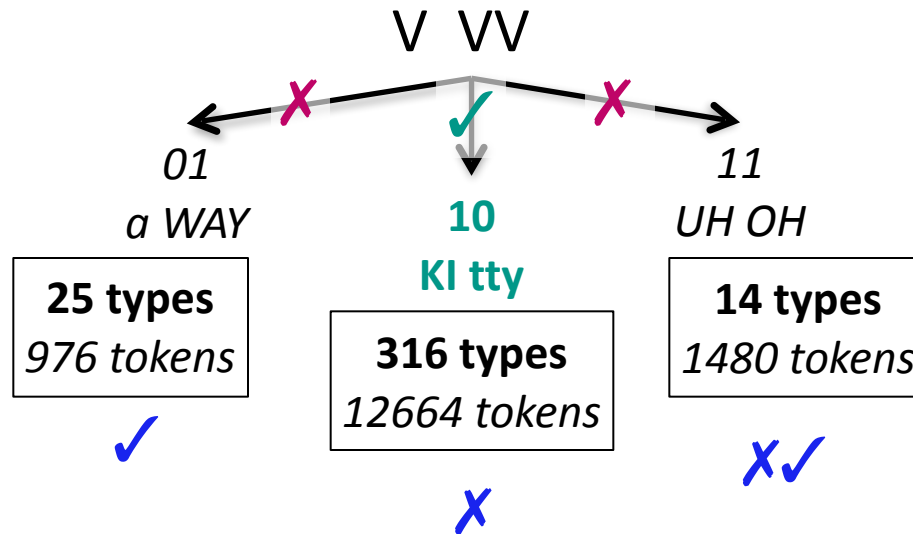
How many exceptions are allowed?  $355 / \ln(355) = 60$

# Productive data filter in action

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

452 syllable word forms

8 syllable distinctions



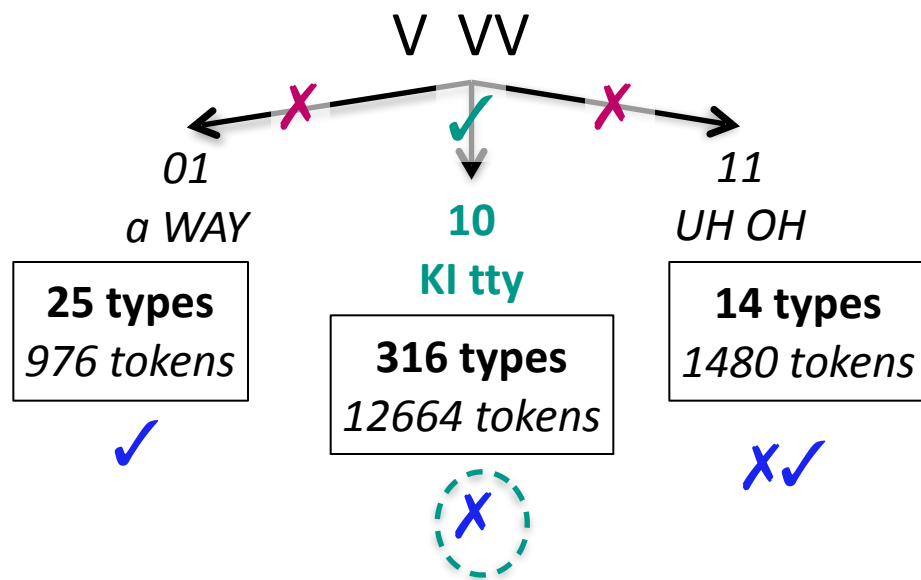
Under the OT syllable representation, there is a dominant stress pattern for this word form. Therefore, this pattern should be accounted for by the English grammar.

# Productive data filter in action

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

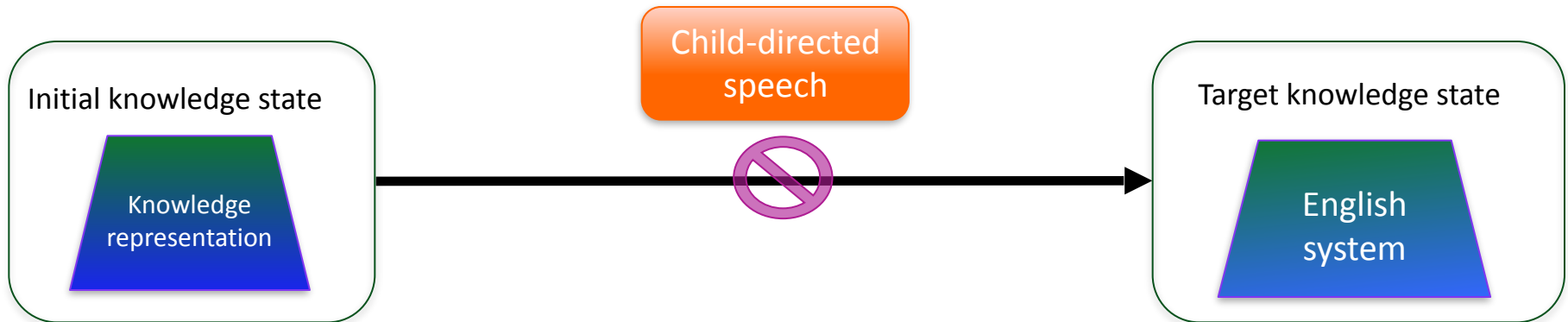
452 syllable word forms

8 syllable distinctions



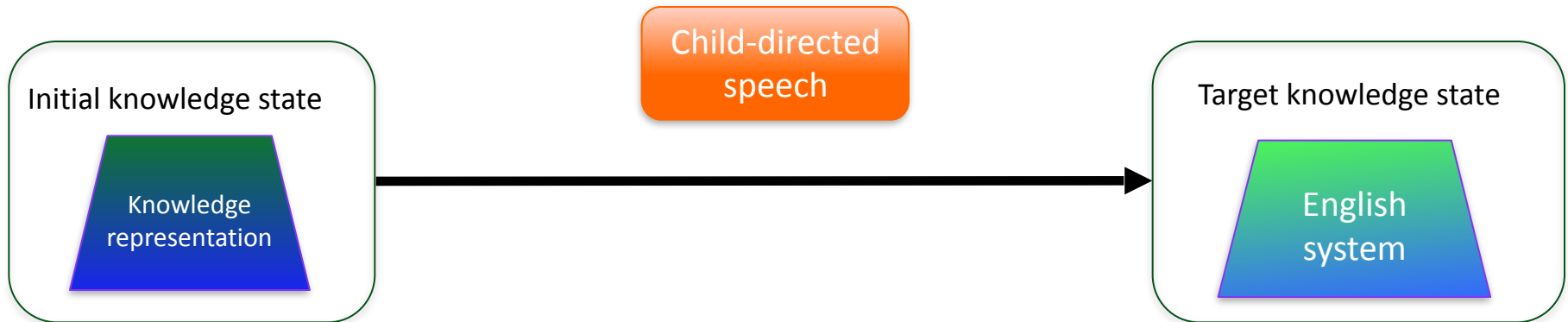
Unfortunately, this is the only pattern the English grammars cannot account for....this means a learner using the productivity filter would have even more trouble learning the current English OT grammar constraints.

# The learnability problem: A third option





# The learnability problem: A third option

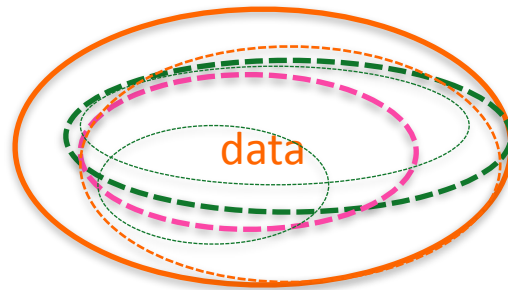


The **target grammar is different** than we think, and we should update our definition of the English grammar.

# What should the target grammar be?

English  
system

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

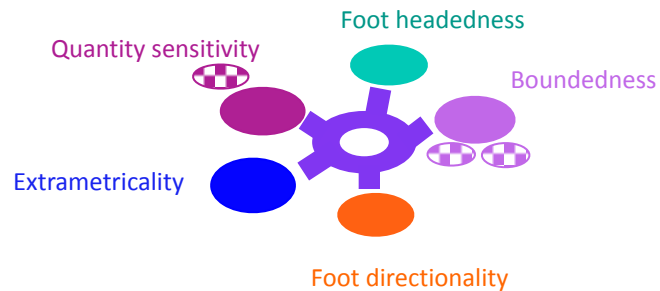


# What should the target grammar be?

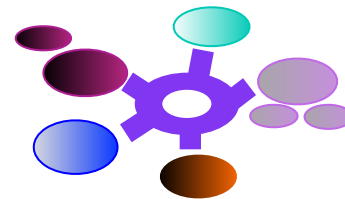
English  
system

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

## Parametric: HV



## English grammar

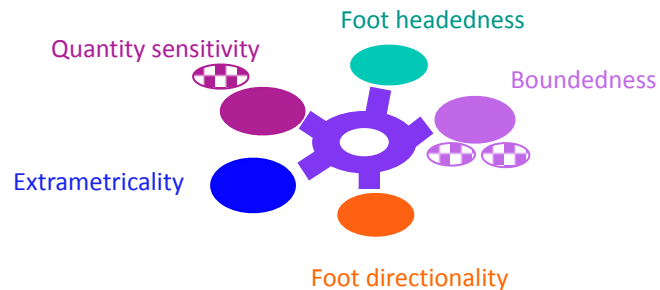


# What should the target grammar be?

English  
system

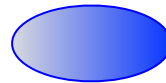
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## Parametric: HV



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:



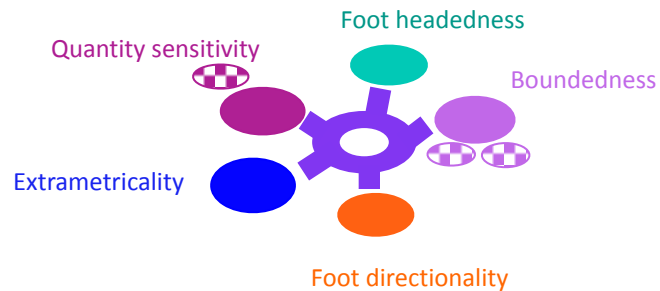
Extrametricality on the rightmost syllable is used by 53 of 58 high compatibility grammars.

# What should the target grammar be?

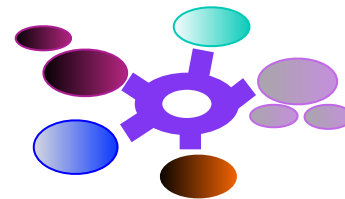
English  
system

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

## Parametric: HV



## English grammar



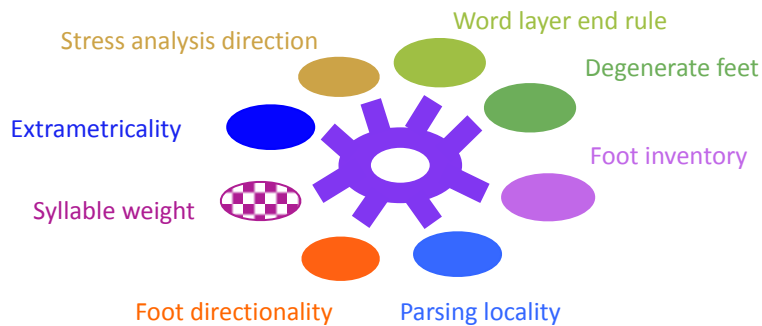
Upshot: Unclear for the HV knowledge representation that the learning problem can be fixed by simply switching one parameter value here or there.

# What should the target grammar be?

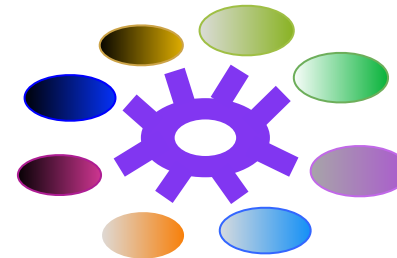
English  
system

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

## Parametric: Hayes



## English grammar

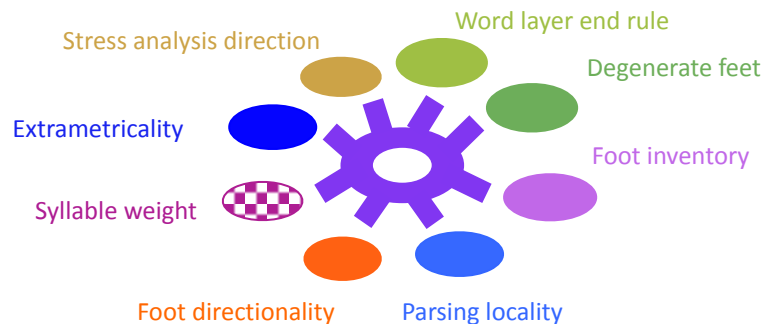


# What should the target grammar be?

English  
system

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

## Parametric: Hayes



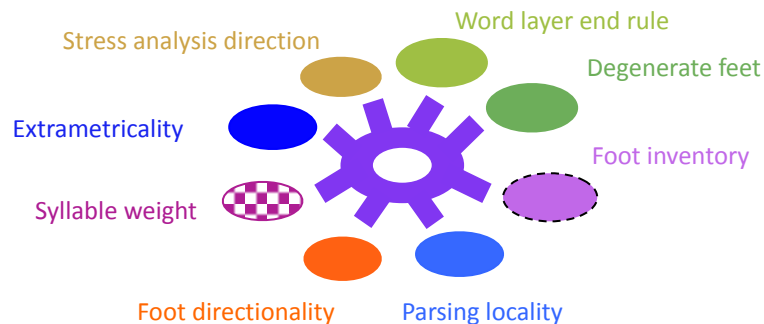
If we examine the grammars with high compatibility, it turns out that there are some parameter values that the majority of high compatibility grammars use, but which the English grammar does not use.

# What should the target grammar be?

English  
system

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. 

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

Optimal: **0.683** types (0.750 tokens)

**Relative compatibility = 0.910**

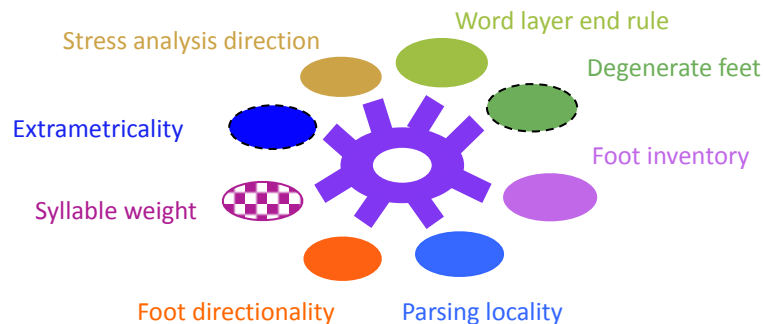


# What should the target grammar be?

English  
system

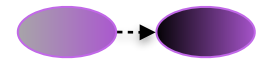
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## 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.

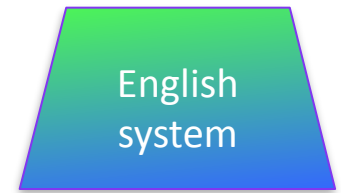


Optimal: **0.683** types (0.750 tokens)

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

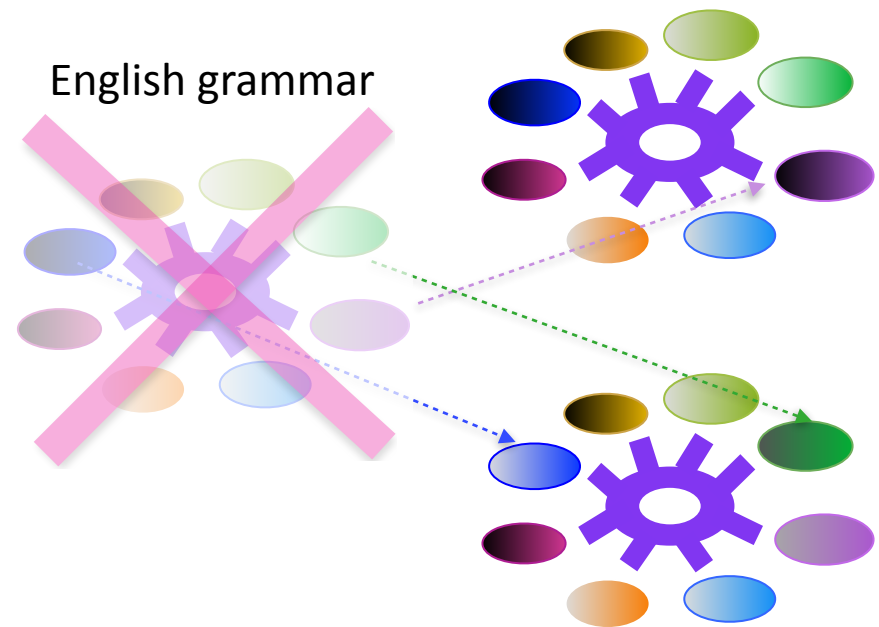
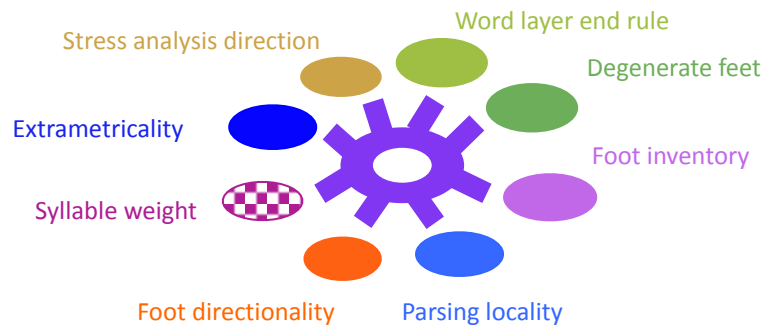
**Relative compatibility = 0.923**

# 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



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?

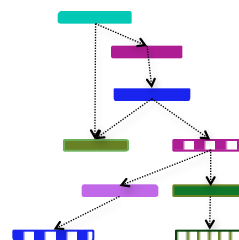
English  
system

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



English grammar obeys this ordering



# What should the target grammar be?

English  
system

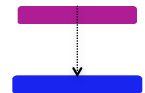
One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

## Constraint-based: OT



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.

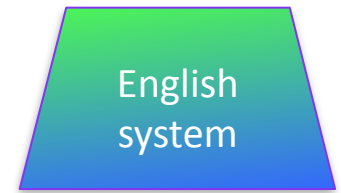
Weight-to-Stress VV >> Nonfinality



(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?

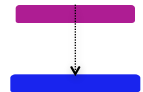


One approach: What values/constraint rankings do the grammars have that are more compatible with the data than the official English grammar?

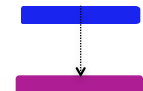
## Constraint-based: OT



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?

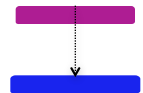
## Constraint-based: OT



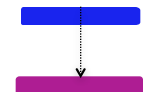
Optimal: 0.657 types (0.729 tokens)

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 raw compatibility with the original ranking is 0.574 types (0.573 tokens).



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



**Relative compatibility = 0.988**

# What should the target grammar be?

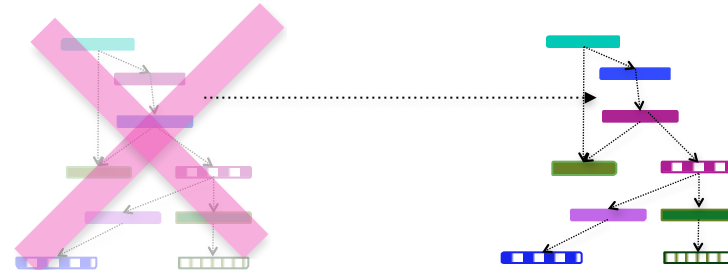
English  
system

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



English grammar obeys this ordering



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

# 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):

<u><b>HV</b></u>		<u><b>Hayes</b></u>		
<b>59.33%</b>	<b>-</b>	<b>48.05%</b>	<b>=</b>	<b>11.28% difference</b>

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. (L = VVC\*, X = V, A = VCC+, P = VC)

<b>Stressed wordform</b>	<b># Types</b>	<b>Examples</b>
<i>Lp</i>	<b>592</b>	<i>water, doing, going</i>
<i>Xp</i>	<b>472</b>	<i>little, getting, coming</i>
<i>Ll</i>	<b>334</b>	<i>baby, sweetie, mommy</i>
<i>Xl</i>	<b>309</b>	<i>kitty, daddy, very</i>
<i>Ap</i>	<b>235</b>	<i>goodness, handsome, helper</i>
<i>LL</i>	<b>188</b>	<i>okay, byebye, TV</i>
<i>Al</i>	<b>172</b>	<i>window, birdie, only</i>
<i>La</i>	<b>171</b>	<i>peanuts, secrets, highest</i>
<i>Xa</i>	<b>170</b>	<i>biggest, buckets, hiccups</i>
<i>xL</i>	<b>145</b>	<i>below, today, hurray</i>

**HV English grammar** derives **8** of the 10 most common stressed wordforms.

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

**Hayes English grammar** derives **5** of the 10 most common stressed wordforms.

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

The **Hayes English grammar** misses **LI and XI**, 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.

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

## HV vs. Hayes on most frequent word forms

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