Why Phonological Learning is Modular

Jeffrey Heinz
heinz@udel.edu

University of Delaware

University of Maryland at College Park
May 6, 2010

Collaborators

James Rogers (Earlham College)

Bill Idsardi (UMCP)

Cesar Koirala, Regine Lai, Darrell Larsen, Tim O'Neill, Jane Chandlee, Robert Wilder, Evan Bradley (University of Delaware)

How can something learn?

1. How do people generalize beyond their experience?
2. How can any*thing* that computes generalize beyond its experience?

- Artificial Intelligence
- Philosophy
- Computer Science
- Linguistics / Language Acquisition
- Psychology
- Natural Language Processing
- ...

Why Phonological Learning is Modular

1. Typological Evidence
2. Formal Learning Theories

> The hypothesis that phonological learning is modular currently offers the best explanation not only for how phonological patterns are learned but also for the character of the typology.

Why Phonological Learning is Modular

1. Typological Evidence
2. Formal Learning Theories

> The hypothesis that phonological learning is modular currently offers the best explanation not only for how phonological patterns are learned but also for the character of the typology.

- Not all the empirical evidence is in yet.

Phonotactics - Knowledge of word well-formedness

ptak thole hlad plast sram mgla vlas flitch dnom rtut

Halle, M. 1978. In *Linguistic Theory and Pyschological Reality*. MIT Press.

Phonotactics - Knowledge of word well-formedness

possible English words	impossible English words
thole	ptak
plast	hlad
flitch	sram
	mgla
	vlas
	dnom
	rtut

1. Question: How do English speakers know which of these words belong to different columns?

Phonotactics - Knowledge of word well-formedness

possible English words	impossible English words
thole	ptak
plast	hlad
flitch	sram
	mgla
	vlas
	dnom
	rtut

1. Question: How do English speakers know which of these words belong to different columns?

Phonotactics - Knowledge of word well-formedness Chumash Version

shtoyonowonowash

stoyonowonowash

stoyonowonowas

shtoyonowonowas

pisotonosikiwat

pisotonoshikiwat

Phonotactics - Knowledge of word well-formedness Chumash Version

possible Chumash words	impossible Chumash words
shtoyonowonowash	stoyonowonowash
stoyonowonowas	shtoyonowonowas
pisotonosikiwat	pisotonoshikiwat

1. Question: How do Chumash speakers know which of these words belong to different columns?
2. By the way, *shtoyonowonowash* means 'it stood upright'

(Applegate 1972)

Phonotactics - Knowledge of word well-formedness Chumash Version

possible Chumash words	impossible Chumash words
shtoyonowonowash	stoyonowonowash
stoyonowonowas	shtoyonowonowas
pisotonosikiwat	pisotonoshikiwat

1. Question: How do Chumash speakers know which of these words belong to different columns?
2. By the way, *shtoyonowonowash* means 'it stood upright'

(Applegate 1972)

Phonotactics - Knowledge of word well-formedness Kwakiutl version

H = syllable with long vowel, L = other syllables

Ĥ	Ĺ	Ĥ L	Ĥ H	L Ĥ
L Ĺ	Ĥ L L	Ĥ L H	Ĥ H L	Ĥ H H
L Ĥ L	L Ĥ H	L L Ĺ	L L Ĥ	L Ĥ L L
L Ĥ L H	Ĥ L L L	Ĥ L L H	Ĥ H L L	Ĥ H L H
L Ĥ H L	L Ĥ H H	Ĥ L H L	Ĥ L H H	Ĥ H H L
Ĥ H H H	L L Ĥ L	L L Ĥ H	L L L Ĺ	L L L Ĥ

Phonotactics - Knowledge of word well-formedness Kwakiutl Version

possible Kwakiutl words				
Ĥ	Ĺ	Ĥ L	Ĥ H	L Ĥ
L Ĺ	Ĥ L L	Ĥ L H	Ĥ H L	Ĥ H H
L Ĥ L	L Ĥ H	L L Ĺ	L L Ĥ	L Ĥ L L
L Ĥ L H	Ĥ L L L	Ĥ L L H	Ĥ H L L	Ĥ H L H
L Ĥ H L	L Ĥ H H	Ĥ L H L	Ĥ L H H	Ĥ H H L
Ĥ H H H	L L Ĥ L	L L Ĥ H	L L L Ĺ	L L L Ĥ

impossible Kwakiutl words
NONE!

1. Question: How do Kwakiutl speakers know this pattern?

Three kinds of phonological patterns

1. **Local sound patterns**; e.g. consonant clusters
   - *#vl, *#pt, . . .
   - Every known language
   - (Chomsky and Halle 1968, many others before and after)
2. **Long-distance sound patterns**; e.g. consonantal and vowel harmony
   - *s. . . sh, . . .
   - Sarcee, Navajo, Finnish, . . .
   - (Hansson 2001, Rose and Walker 2004, Ringen 1988, Baković 2000, Finley 2008, and many others)
3. **Stress patterns over syllables**
   - Every odd syllable, Leftmost heavy otherwise rightmost
   - Pirahã, Pintupi, . . .
   - (Hyman 1977, Halle and Vergnaud 1987, Idsardi 1992, Hayes 1995, Hyde 2001, Gordon 2002, Goedemans 2005, van der Hulst 2009, Heinz 2009, and many others)

Limits on the variation

1. **Local sound patterns**; e.g. consonant clusters
   - ?
2. **Long-distance sound patterns**; e.g. consonantal and vowel harmony
   - Consonantal harmony patterns do not exhibit blocking: e.g. *s...sh unless [z] intervenes.
     (Hansson 2001, Rose and Walker 2004)
   - No harmony pattern applies only to the first and last sounds.
3. **Stress patterns over syllables**
   - The middle syllable gets a beat (Single)
   - Every fourth syllable gets a beat (Quaternary)
   - Every fifth syllable gets a beat (Quinary)
   - ...
   - The prime-numbered syllables (2,3,5,7,11,...) get a beat
   - The prime-numbered syllables minus one (1,2,4,6,10,...) get a beat
   - ...

Computational Theory: Three Important Questions

1. Does it exist?
2. Is it computable?
3. Is it feasibly computable?

Formal Language Theory

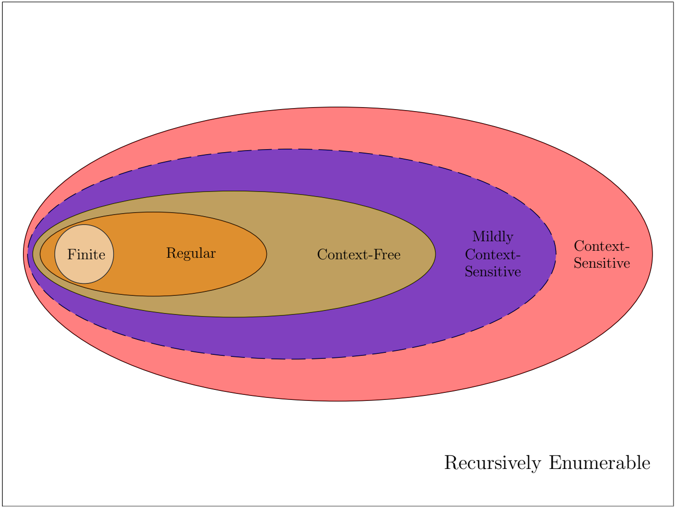


Figure: The Chomsky hierarchy classifies logically possible patterns.

Chomsky 1956, 1959, Harrison 1978

Formal Language Theory

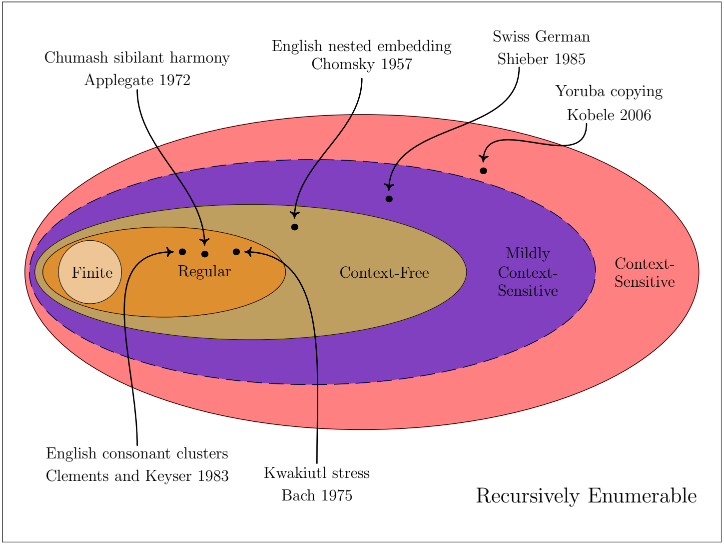


Figure: Natural language patterns in the Chomsky hierarchy.

Formal Language Theory

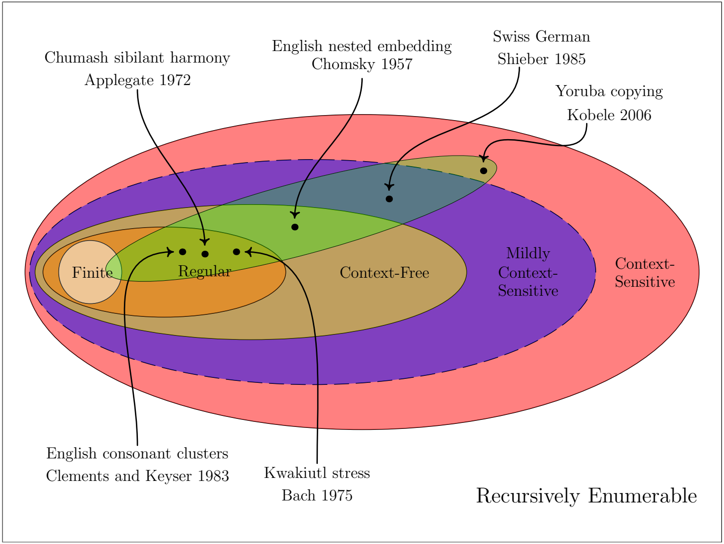


Figure: Possible theories of natural language.

Formal Language Theory

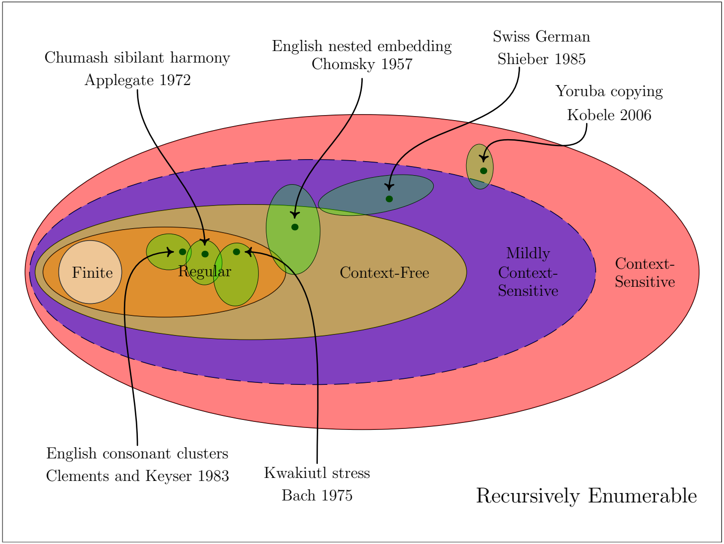


Figure: Possible theories of natural language.

Formal Learning Theory

1. How can we define “learning”?
2. Under the definition, what can be learned and how?

Formal Learning Theory

1. How can we define "learning"?
2. Under the definition, what can be learned and how?

Learning requires a structured hypothesis space, which excludes at least some finite-list hypotheses.

Gleitman 1990, p. 12:

> *'The trouble is that an observer who notices everything can learn nothing for there is no end of categories known and constructable to describe a situation [emphasis in original].'*

Formal Learning Theories

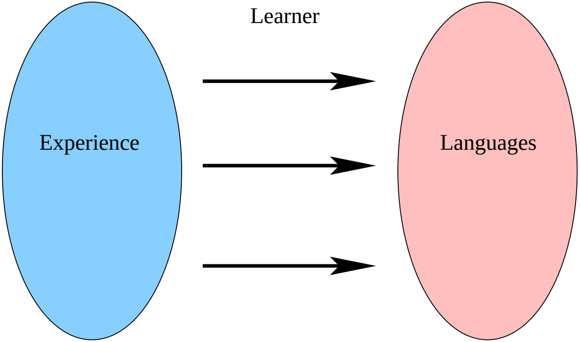


Figure: Learners are functions ϕ from experience to languages.

(Gold 1967, Horning 1969, Angluin 1980, Osherson et al. 1984, Angluin 1988, Anthong and Biggs 1991, Kearns and Vazirani 1994, Vapnik 1994, 1998, Jain et al. 1999, Niyogi 2006, de la Higuera 2010)

The Experience

1. It is a sequence.
2. It is finite.

$$\begin{array}{l} w_0 \\ w_1 \\ w_2 \\ \dots \\ w_n \end{array} \qquad \downarrow \text{time}$$

Types of Experience

1. Positive evidence
2. Positive and negative evidence
3. Noisy evidence
4. Queried Evidence

$$\begin{array}{c} w_0 \in L \\ w_1 \in L \\ w_2 \in L \\ \ldots \\ w_n \in L \end{array} \qquad \Big\downarrow \text{time}$$

Types of Experience

1. Positive evidence
2. Positive and negative evidence
3. Noisy evidence
4. Queried Evidence

$$\begin{array}{c} w_0 \in L \\ w_1 \notin L \\ w_2 \notin L \\ \dots \\ w_n \in L \end{array} \qquad \Big\downarrow \text{time}$$

Types of Experience

1. Positive evidence
2. Positive and negative evidence
3. Noisy evidence
4. Queried Evidence

$$\begin{array}{c} w_0 \in L \\ w_1 \notin L \\ w_2 \in L \ (\text{but in fact } w_2 \notin L) \\ \dots \\ w_n \in L \end{array} \qquad \downarrow \text{time}$$

Types of Experience

1. Positive evidence
2. Positive and negative evidence
3. Noisy evidence
4. Queried Evidence

$$w_0 \in L$$
$$w_1 \notin L$$
$w_2 \in L$ (because learner specifically asked about w_2)
$$\dots$$
$$w_n \in L$$

$\downarrow$ time

The Languages

1. They can be sets of words or distributions over words.
2. They are computable.

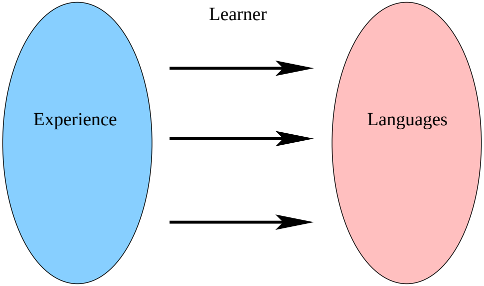


Figure: Learners are functions ϕ from experience to languages.

The Languages

1. They can be sets of words or distributions over words.
2. They are computable.

   I.e. they are describable with grammars.

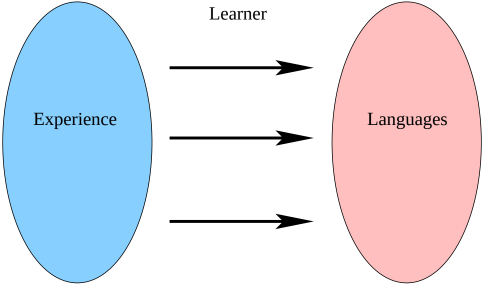


Figure: Learners are functions ϕ from experience to languages.

The Languages

1. They can be sets of words or distributions over words.
2. They are computable.

   I.e. they are describable with grammars.

   I.e they are r.e. languages.

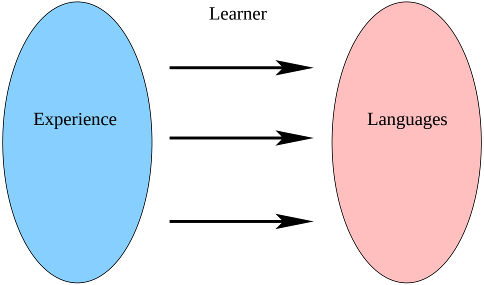


Figure: Learners are functions ϕ from experience to languages.

The Languages

1. They can be sets of words or distributions over words.
2. They are computable.

   I.e. they are describable with grammars.

   I.e they are r.e. languages.

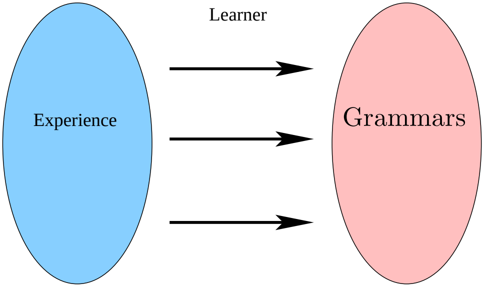


Figure: Learners are functions ϕ from experience to grammars.

Learning Criteria

1. What does it mean to learn a language?
2. What kind of experience is required for success?
3. What counts as success?

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$
w_2	$\phi(\langle w_0, w_1, w_2 \rangle) = G_2$

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$
w_2	$\phi(\langle w_0, w_1, w_2 \rangle) = G_2$
$\ldots$	

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$
w_2	$\phi(\langle w_0, w_1, w_2 \rangle) = G_2$
$\ldots$	
w_n	$\phi(\langle w_0, w_1, w_2, \ldots, w_n \rangle) = G_n$

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$
w_2	$\phi(\langle w_0, w_1, w_2 \rangle) = G_2$
$\ldots$	
w_n	$\phi(\langle w_0, w_1, w_2, \ldots, w_n \rangle) = G_n$
$\ldots$	

$\downarrow$ time

What does it mean to learn a language?

1. Convergence.
2. Imagine an infinite sequence. Is there some point n after which the learner's hypothesis doesn't change (much)?

datum	Learner's Hypothesis
w_0	$\phi(\langle w_0 \rangle) = G_0$
w_1	$\phi(\langle w_0, w_1 \rangle) = G_1$
w_2	$\phi(\langle w_0, w_1, w_2 \rangle) = G_2$
$\dots$	
w_n	$\phi(\langle w_0, w_1, w_2, \dots, w_n \rangle) = G_n$
$\dots$	
w_m	$\phi(\langle w_0, w_1, w_2, \dots, w_m \rangle) = G_m$

$\downarrow$ time

Does $G_m \simeq G_n$?

What kind of experience is required for success?

Types of Experience

1. Positive-only or positive and negative evidence.
2. Noisless or noisy evidence.
3. Queries allowed or not?

Which infinite sequences require convergence?

1. only complete ones? I.e. where every piece of information occurs at some finite point
2. only computable ones? I.e. the infinite sequence itself is describable by some grammar

What kind of experience is required for success?

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

What kind of experience is required for success?

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

1. Identification in the limit from positive data (Gold 1967)

What kind of experience is required for success?

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

2. Identification in the limit from positive and negative data

(Gold 1967)

What kind of experience is required for success?

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

3. Identification in the limit from positive data from r.e. texts
(Gold 1967)

4. Learning context-free and r.e. distributions
(Horning 1969, Angluin 1988)

What kind of experience is required for success?

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

5. Probably Approximately Correct learning
(Valiant 1984, Anthony and Biggs 1991, Kearns and Vazirani 1994

What counts as success?

We are interested in learners of *classes of languages* and not just a single language.

Why?

What counts as success?

We are interested in learners of *classes of languages* and not just a single language.

Why?

Because every language can be learned by a constant function!

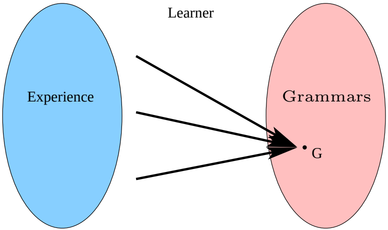
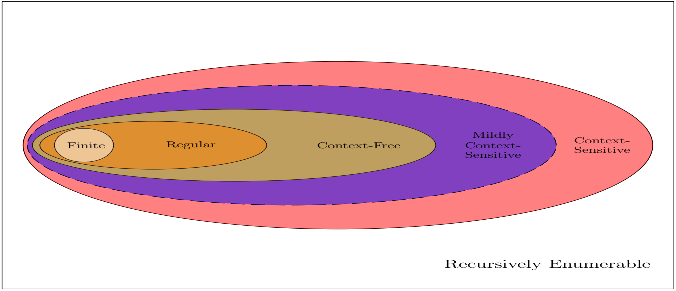


Figure: Learners are functions ϕ from experience to grammars.

Results of Formal Learning Theories: Existence

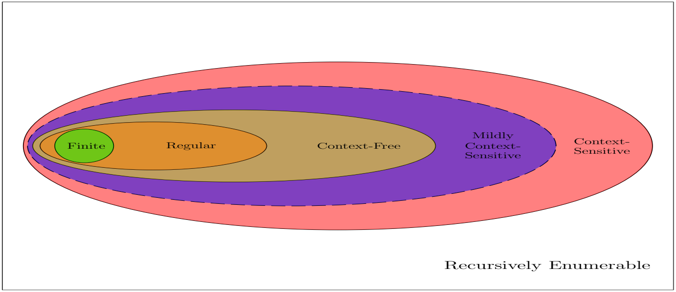
Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence



Results of Formal Learning Theories: Existence

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

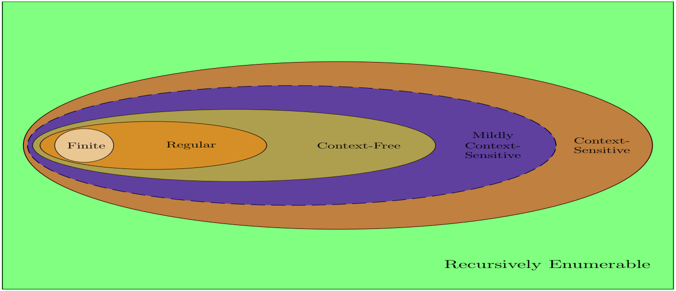
1. Identification in the limit from positive data (Gold 1967)



Results of Formal Learning Theories: Existence

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

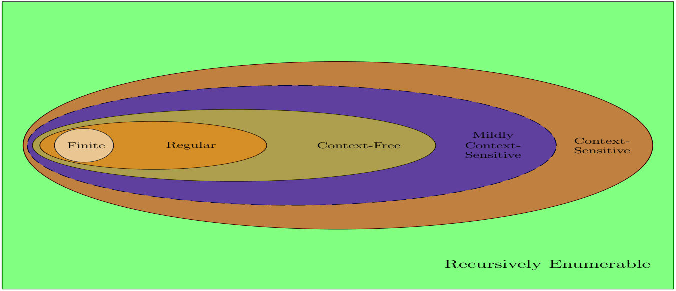
2. Identification in the limit from positive and negative data (Gold 1967)



Results of Formal Learning Theories: Existence

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

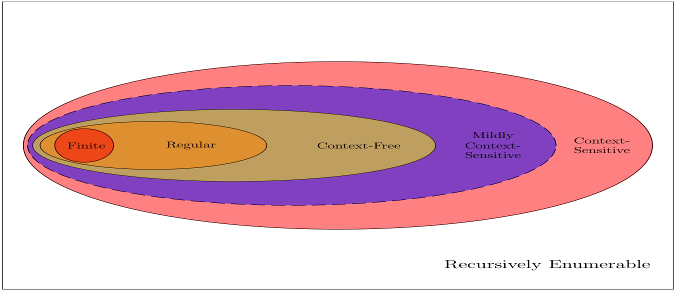
3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988)



Results of Formal Learning Theories: Existence

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

5. Probably Approximately Correct learning
(Valiant 1984, Anthony and Biggs 1991, Kearns and Vazirani 1994



Results of Formal Learning Theory: Feasibility

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

Results of Formal Learning Theory: Feasibility

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

1. Identification in the limit from positive data (Gold 1967)

   No superfinite class is learnable.
   The finite class is feasibly learnable.

Results of Formal Learning Theory: Feasibility

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

2. Identification in the limit from positive and negative data (Gold 1967)

   The r.e. class is learnable but NOT even the regular class is feasibly learnable.

Results of Formal Learning Theory: Feasibility

Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

3. Identification in the limit from positive data from r.e. texts (Gold 1967)
4. Learning context-free and r.e. distributions (Horning 1969, Angluin 1988)

   The r.e. class of languages and distributions is learnable but NOT even the regular class is feasibly learnable.

Results of Formal Learning Theory: Feasibility

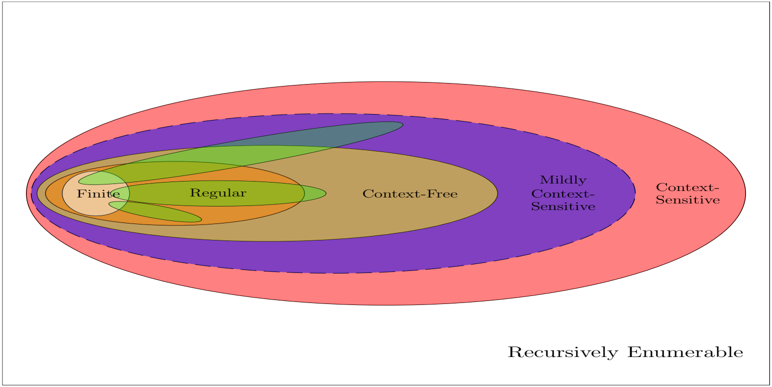
Makes learning easier	Makes learning harder
positive and negative evidence	positive evidence only
noiseless evidence	noisy evidence
queries permitted	queries not permitted
approximate convergence	exact convergence
complete infinite sequences	any infinite sequence
computable infinite sequences	any infinite sequence

5. Probably Approximately Correct learning
(Valiant 1984, Anthony and Biggs 1991, Kearns and Vazirani 1994)

Not even the finite class of languages is learnable.

Formal Learning Theory: Positive Results

Many classes which cross-cut the Chomsky hierarchy and exclude some finite languages are feasibly learnable in the senses discussed.



(Angluin 1980, 1982, Garcia et al. 1990, Muggleton 1990, Denis et al. 2002, Fernau 2003, Yokomori 2003, Oates et al. 2006, Niyogi 2006, Clark and Eryaud 2007, Heinz 2008, to appear, Yoshinaka 2008, Case et al. 2009, de la Higuera 2010)

Summary

1. Structured, restricted hypothesis spaces can be feasibly learned.
2. The positive learning results are proven results, and the proofs are often constructive.
3. The claim that "statistical learning" is more powerful than "symbolic learning" mischaracterizes the learning issues.
4. The real issue is whether or not success ought to be defined only with respect to data sequences generable by fixed, unchanging distributions (e.g. computable ones).

Putting it all together

1. I am not claiming the following learners are the full story.
2. I am claiming that they are good approximations to the full story and that the full story will incorporate their key elements.
3. The role of phonological features, similarity, sonority, etc. is ongoing and will refine the present proposals.

Local sound patterns

Distinctions are made on the basis of contiguous subsequences.

possible English words	impossible English words
thole	ptak
plast	hlad
flitch	sram
	mgla
	vlas
	dnom
	rtut

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip ✓

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip ✓

ptip

Local sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Local (McNaughton and Papert 1971, Rogers and Pullum 2007)
2. They are subregular and exclude some finite languages.
3. If every k-long contiguous subsequence is licensed by the grammar, the word belongs to the language.

stip ✓

ptip $\times$

Learning local sound patterns

1. Stricly k-Local languages are identifiable in the limit from positive data (Garcia et al. 1990).
2. Stricly k-Local distributions can be efficiently estimated (Jurafsky & Martin 2008) (they are n-gram models)
3. **Keep track of the observed k-long contiguous subsequences.**

i	$t(i)$	$SL_2(t(i))$	Grammar G	$L(G)$
-1			$\emptyset$	$\emptyset$
0	$aaaa$	$\{aa\}$	$\{\mathbf{aa}\}$	aaa^*
1	aab	$\{aa,\ ab\}$	$\{aa,\ \mathbf{ab}\}$	$aaa^* \ \cup \ aaa^*b$
2	ba	$\{ba\}$	$\{aa,\ ab,\ \mathbf{ba}\}$	$\Sigma^*/\Sigma^*bb\Sigma^*$
...				

The Strictly 2-Local learner learns *bb

Long-distance sound patterns

Distinctions are made on the basis of potentially discontiguous subsequences.

possible Chumash words	impossible Chumash words
shtoyonowonowash	stoyonowonowash
stoyonowonowas	shtoyonowonowas
pisotonosikiwat	pisotonoshikiwat

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos ✓

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos ✓

sotosh

Long-distance sound patterns and formal language theory

1. The formal languages which make distinctions on the basis of k-long contiguous subsequences are called Strictly k-Piecewise (Heinz 2007, Rogers et al. 2009, Heinz to appear).
2. They are subregular and exclude some finite languages.
3. If every k-long subsequence is licensed by the grammar, the word belongs to the language.

sotos ✓

sotosh ×

Learning long-distance sound patterns

1. Stricly k-Piecewise languages are identifiable in the limit from positive data (Heinz 2007, to appear).
2. Stricly k-Piecewise distributions can be efficiently estimated (Heinz & Rogers to appear)
3. **Keep track of the observed k-long subsequences.**

i	$t(i)$	$SP_2(t(i))$	Grammar G	Language of G
-1			$\emptyset$	$\emptyset$
0	$aaaa$	$\{\lambda, a, aa\}$	{λ, **a, aa**}	a^*
1	aab	$\{\lambda, a, b, aa, ab\}$	{λ, a, aa, **b, ab**}	$a^* \cup a^*b$
2	baa	$\{\lambda, a, b, aa, ba\}$	{λ, a, b, aa, ab, **ba**}	$\Sigma^* \backslash (\Sigma^* b \Sigma^* b \Sigma^*)$
3	aba	$\{\lambda, a, b, ab, ba\}$	{λ, a, b, aa, ab, ba}	$\Sigma^* \backslash (\Sigma^* b \Sigma^* b \Sigma^*)$
...				

The learner ϕ_{SP_2} learns *b...b

Further comments

1. Like the regions in the Chomsky hierarchy, the Strictly Local and Strictly Piecewise classes have multiple, independent, converging characterizarions from formal language theory, automata theory, and logic.
2. They are incomparable.
3. Consequently, Strictly Local learners cannot learn Strictly Piecewise patterns and vice versa.
4. Strictly Piecewise learners cannot learn:
   - blocking patterns, e.g. *s…sh unless [z] intervenes.
   - harmony patterns which apply only to the first and last sounds.

Stress Patterns

1. Combining two typological studies (Bailey 1995 and Gordon 2002) yields a survey of 405 languages (423 descriptions and 109 distinct patterns).
2. None are Strictly Piecewise for any k.
3. At least 19 are not Strictly Local for any k. (e.g. Kwakiutl)
4. All but 2 (somewhat controversial cases) are **neighborhood-distinct** (Heinz 2009).

Learning Stress Patterns

1. **Neighborhood-distinctness** is a locality condition.
2. A learner which uses this property is able to identify in the limit 100 of the 109 distinct stress patterns and get awfully close to the other 9 (Heinz 2009).
3. None of the following patterns are neighborhood-distinct nor learnable by the neighborhood-distinct learner.
    - The middle syllable gets a beat (Single)
    - Every fourth syllable gets a beat (Quaternary)
    - Every fifth syllable gets a beat (Quinary)
    - ...
    - The prime-numbered syllables (2,3,5,7,11,...) get a beat
    - The prime-numbered syllables minus one (1,2,4,6,10,...) get a beat
    - ...

Summary

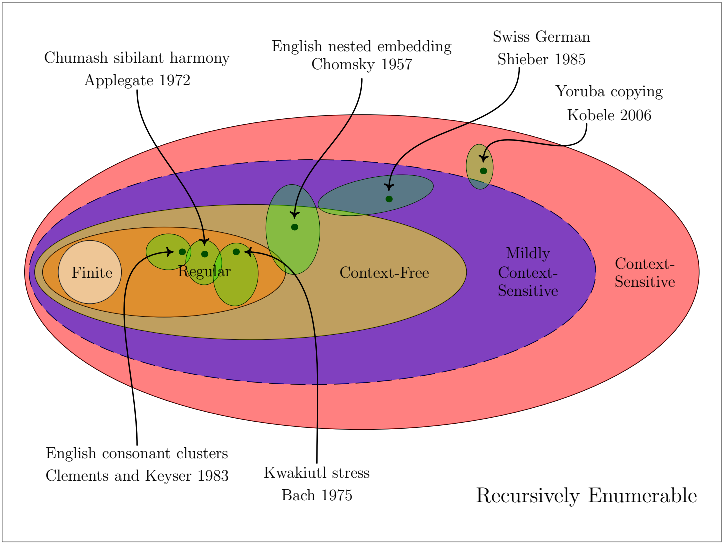


Figure: Strictly Local, Strictly Piecewise and Neighborhood-distinct classes.

Summary

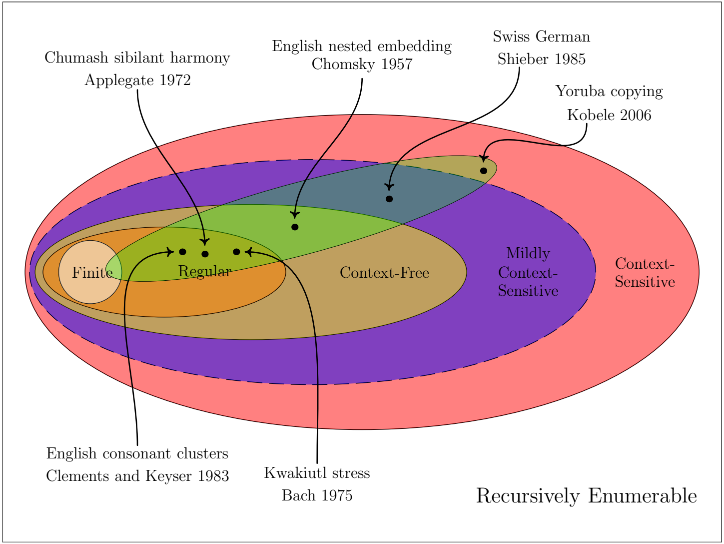


Figure: Where is the learner of this class?

Modular Learning and Biology

Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms...

From a biological perspective, the idea of a general-learning mechanism is equivalent to assuming that there is a general-purpose sensory organ, which solves the problem of sensing.

(Gallistel and King 2009:218)

Artificial language learning experiments

1. Can people learn unattested logically possible patterns?
2. Do people generalize the same way within different linguistic domains?

Preliminary Experimental Results

- Subjects appear to learn consonantal harmony with blocking (Samuels, in progress)
- If the same formal pattern is present in segmental patterns and in stress patterns (over syllables), subjects generalize differently (Bergelson et. al 2010).
- Much additional work in progress

Conclusion

1. Linguistic patterns are not arbitrary.
2. Only structured classes of patterns can be learned.
3. Distinct, feasible learning models for distinct phonological patterns exist.
4. These help explain the character of the typology.
5. A single, feasible learning model for these distinct phonological patterns does not exist (yet, ever?).
6. Such a model is likely to have to attribute the character of the typology to something else.
7. Artificial language learning experiments can help.

Conclusion

1. Linguistic patterns are not arbitrary.
2. Only structured classes of patterns can be learned.
3. Distinct, feasible learning models for distinct phonological patterns exist.
4. These help explain the character of the typology.
5. A single, feasible learning model for these distinct phonological patterns does not exist (yet, ever?).
6. Such a model is likely to have to attribute the character of the typology to something else.
7. Artificial language learning experiments can help.

The hypothesis that phonological learning is modular currently offers the best explanation not only for how phonological patterns are learned but also for the character of the typology.

Thank you