Computation, Phonology, and Typology

Jeffrey Heinz
heinz@udel.edu

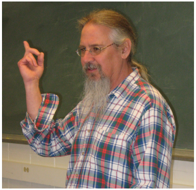
University of Delaware

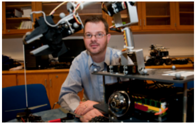
StonyBrook University
May 2, 2014

*This research has received support from NSF awards CPS#1035577 and LING#1123692.

Collaborators

- Prof. Jim Rogers (Earlham College)
- Prof. Herbert G. Tanner (UD)
- Prof. Harry van der Hulst (UConn)
- Prof. Bill Idsardi (UMCP)
- Prof. Rémi Eryaud (Marseilles)
- Dr. Jane Chandlee, PhD 2014
- Dr. Regine YeeKing Lai, PhD 2012
- Cesar Koirala (PhD exp. 2014)
- Adam Jardine (PhD exp. 2016)
- Amanda Payne (PhD exp. 2016)
- Hyun Jin Hwangbo (PhD exp. 2017)
- Huan Luo (PhD exp. 2017)
- Brian Gainor (LDC)
- Sean Wibel (U. Washington)


Jim


Bert


Regine


Cesar


Adam


Amanda

Unpictured
Harry, Bill, Rémi, Jane, Hyun Jin, Huan, Brian, Sean

Wilhelm Von Humboldt

"language makes infinite use of finite means"

Wilhelm Von Humboldt

Typology:

1. “Encyclopedia of Types”
2. “Encyclopedia of Categories”

What is phonology?

A point of agreement between different theories of phonology

- There exist underlying representations of morphemes which are mapped to surface representations.

Fundamental questions of phonological theory

1. What is the nature of the abstract, lexical (‘underlying’) representations?
2. What is the nature of the surface forms?
3. What is the nature of the mapping from underlying forms to surface forms?

What is phonology?

A point of agreement between different theories of phonology

- There exist underlying representations of morphemes which are mapped to surface representations.

Fundamental questions of phonological theory

1. What is the nature of the abstract, lexical (‘underlying’) representations?
2. What is the nature of the surface forms?
3. What is the nature of the mapping from underlying forms to surface forms?

What is phonology?

A point of agreement between different theories of phonology

- There exist underlying representations of morphemes which are mapped to surface representations.

Fundamental questions of phonological theory

1. What is the nature of the abstract, lexical ('underlying') representations?
2. What is the nature of the surface forms?
3. What is the nature of the mapping from underlying forms to surface forms?

The ‘encyclopedias’ in this talk

Encyclopedia of Types

- Surveys of phonotactic patterns
- Surveys of phonological mappings

Encyclopedia of Categories

- Computer Science
- Specifically: a model theoretic approach to formal language theory (Rogers 1994, Graf 2010)

Broader goals of this talk

1. Explain why computational characterizations of language patterns matter
   - help identify abstract universal properties of natural language
   - help identify inductive principles which can explain how natural language patterns are learned
2. Explain the subregular computational classes that phonological generalizations appear to belong to.

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

Phonotactics - Knowledge of word well-formedness

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

This knowledge can be modeled as a stringset

Example

All possible English words are in the set; all logically possible, impossible words are out of the set.

This knowledge can be modeled as a stringset

Example

All possible English words are in the set; all logically possible, impossible words are out of the set.

$$mgl$$

This knowledge can be modeled as a stringset

Example

All possible English words are in the set; all logically possible, impossible words are out of the set.

$$mgl \cdot \Sigma^*$$

This knowledge can be modeled as a stringset

Example

All possible English words are in the set; all logically possible, impossible words are out of the set.

$$\overline{mgl \cdot \Sigma^*}$$

This knowledge can be modeled as a stringset

Example

All possible English words are in the set; all logically possible, impossible words are out of the set.

$$\overline{mgl \cdot \Sigma^*} \cap \overline{pt \cdot \Sigma^*} \cap \ldots$$

This knowledge can be modeled as a stringset

Example

Any markedness constraint in Optimality Theory.

All surface forms with zero violations are in the set; all surface forms with nonzero violations are out of the set.

Mappings can be modeled as sets of pairs (relations)

Word-final obstruent devoicing

$$[\text{-sonorant}] \longrightarrow [\text{-voice}]\ /\ \_\#$$

$$*[+\textsc{voice},\text{-}\textsc{sonorant}]\#,\ \textsc{Max-C} >> \text{ID}(\textsc{voice})$$

(rat, rat)	(sap, sap)
(rad, rat)	(sab, sap)
...	(sag, sat)
(flugenrat, flugenrat)	...
(flugenrad, flugenrat)	
...	

Objects of Linguistic Inquiry

1. These infinite sets of strings and infinite sets of pairs are an object of linguistic inquiry.
2. Grammars are one way to describe them.
3. A theory of grammars speaks to what is a possible generalization (stringset or mapping).

How can we compare the phonologies of different languages?

How can we compare the phonologies of different languages?

Inventories

We can measure the size of the phonemic inventory.
(Maddieson 1984, 1992, et seq. . . . Atkinson 2011)

How can we compare the phonologies of different languages?

But what about phonological processes or constraints?

Constraints and processes describe sets of strings and mappings from one set to another. These objects are of *infinite* size so counting doesn't help!

How can we compare the phonologies of different languages?

Measure the size of grammars.

1. SPE. Size of rules (feature counting)
2. Principles and Parameters. Number of parameters to set.
3. OT. Count "relevant" constraints/rankings if they are innate (T-orders (Antilla 2008); r-volume (Riggle))

How can we compare the phonologies of different languages?

Computational complexity.

There exist independently-motivated, converging mathematical criteria for ordering the complexity of these infinite objects.

- These characterizations were developed in the early 1970s (McNaughton and Papert 1971), but were not applied to linguistic theory until the 1990s.
- These criteria have been argued to be important cognitively (Rogers and Pullum 2011, Rogers et al. 2013, Heinz and Idsardi 2013).

Classifying Sets of Strings

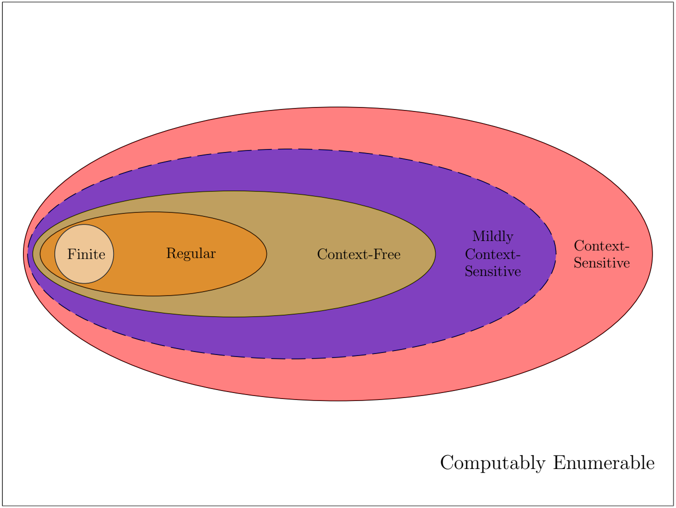
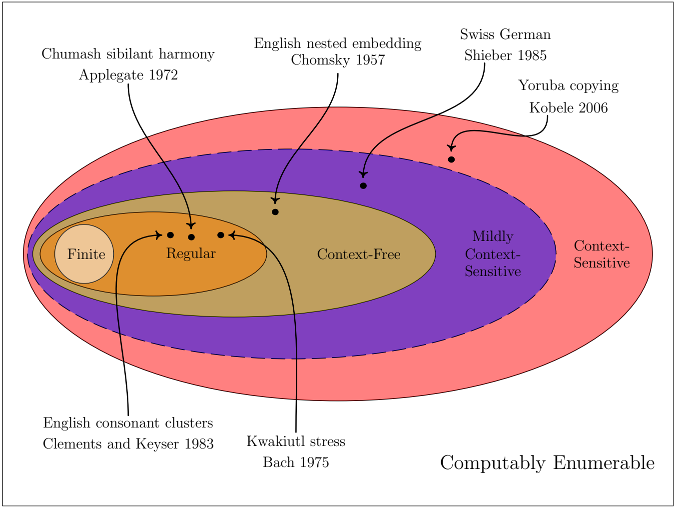


Figure: The Chomsky hierarchy

computably enumerable
|
context-sensitive
|
mildly context-sensitive
|
context-free
|
regular
|
finite

Classifying Sets of Strings



computably enumerable
|
context-sensitive
|
mildly context-sensitive
|
context-free
|
regular
|
finite

Figure: Natural language patterns in the hierarchy.

Phonological mappings are regular

(Johnson 1972, Koskenniemi 1983, Kaplan and Kay 1994)

1. Optional, left-to-right, right-to-left, and simultaneous application of SPE-style rules $A \longrightarrow B \ / \ C \,\_\_ D$ (where A,B,C,D are regular sets) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition.
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all attested phonological patterns.

Phonological mappings are regular

(Johnson 1972, Koskenniemi 1983, Kaplan and Kay 1994)

1. Optional, left-to-right, right-to-left, and simultaneous application of SPE-style rules $\text{A} \longrightarrow \text{B} \ / \ \text{C} \ \_\text{D}$ (where A,B,C,D are regular sets) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition.
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all attested phonological patterns.

Phonological mappings are regular

(Johnson 1972, Koskenniemi 1983, Kaplan and Kay 1994)

1. Optional, left-to-right, right-to-left, and simultaneous application of SPE-style rules $A \longrightarrow B \ / \ C \_\_ D$ (where A,B,C,D are regular sets) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition.
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all attested phonological patterns.

Phonological mappings are regular

(Johnson 1972, Koskenniemi 1983, Kaplan and Kay 1994)

1. Optional, left-to-right, right-to-left, and simultaneous application of SPE-style rules $A \longrightarrow B \ / \ C \_\_ D$ (where A,B,C,D are regular sets) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition.
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all attested phonological patterns.

Phonological mappings are regular

(Johnson 1972, Koskenniemi 1983, Kaplan and Kay 1994)

1. Optional, left-to-right, right-to-left, and simultaneous application of SPE-style rules $A \longrightarrow B \mathbin{/} C \_\_ D$ (where A,B,C,D are regular sets) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.
2. Rule ordering is functional composition.
3. Regular relations are closed under composition.
4. SPE grammars (finitely many ordered rewrite rules of the above type) can describe virtually all attested phonological patterns.

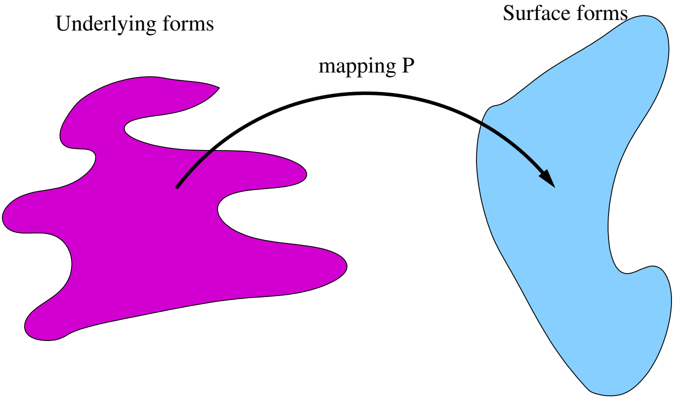
Therefore, phonological mappings are regular relations.

Regardless of whether they are described with SPE, OT, or other formalisms!

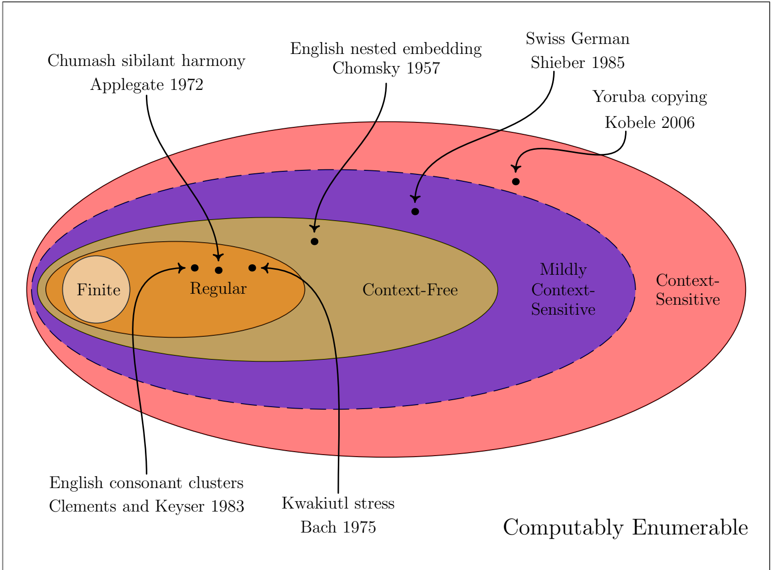
Regular mappings entail *regular* phonotactics and *regular* morpheme structure constraints

Theorem (Rabin and Scott 1959)

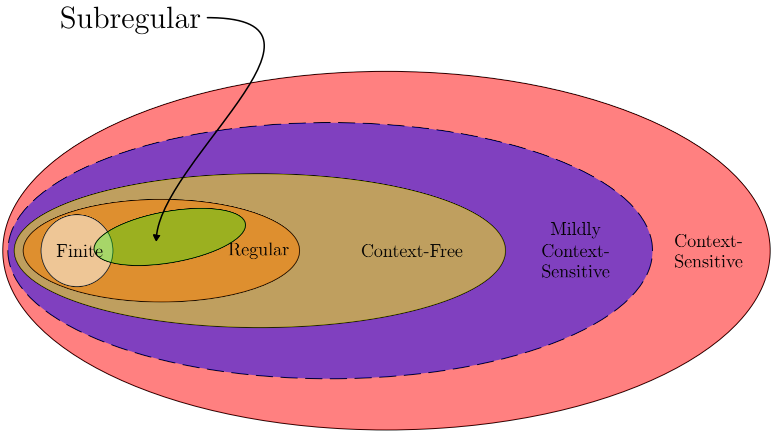
The domain and image of regular relations are regular stringsets.



“Being regular” is a start, but it is not sufficient to make the distinctions we want



"Being regular" is a start, but it is not sufficient to make the distinctions we want



Aside on the Generative Capacity of classic OT

1. (undergeneration) It's reasonably well-known that classic OT (Prince and Smolensky 1993, 2004) cannot describe opaque mappings (Idsardi 1998, 2000, Buccola 2013).
2. (overgeneration) It's less well-known that optimization over simple markedness and faithfulness can result in *nonregular* mappings (Riggle 2004, Gerdemann and Hulden 2012).

Enriching the Encyclopedia of Categories

Better characterizations of phonological patterns

- Allows us to distinguish phonological patterns according to independent measures of complexity
- Leads to stronger universals
- And thus to new hypotheses regarding what a *humanly* possible phonological pattern is

Enriching the Encyclopedia of Categories

Payoffs for better understanding *learning*

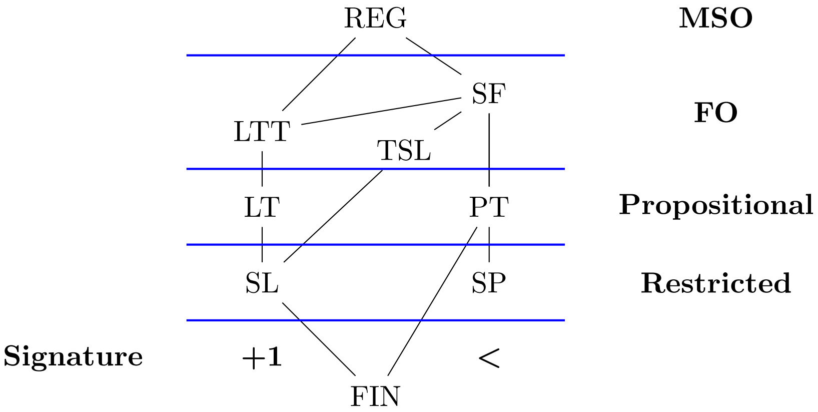
- Are the stronger universals useful for learning?
  (what Moreton 2008 calls analytic bias)

Enriching the Encyclopedia of Categories

Payoffs for natural language processing

- Insights can be incorporated into NLP algorithms
- Factoring and composition may occur with lower complexity

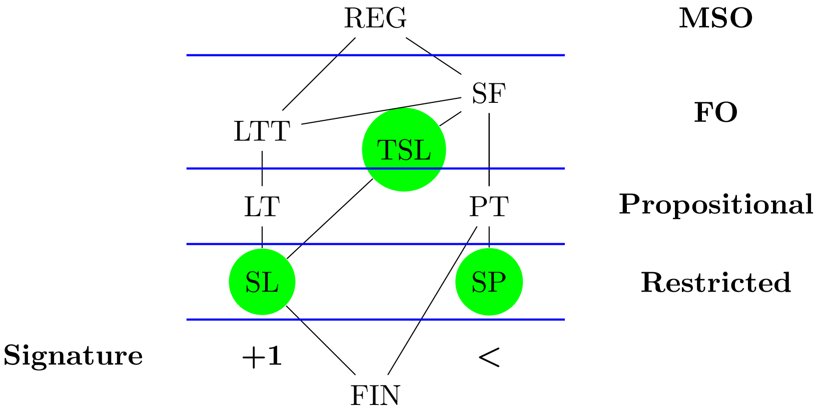
Interesting *subregular* classes of stringsets



(McNaughton and Papert 1971, Rogers et al. 2010, 2013, Heinz et al. 2011)

LTT	Locally Threshold Testable	TSL	Tier-based Strictly Local
LT	Locally Testable	PT	Piecewise Testable
SL	Strictly Local	SP	Strictly Piecewise

Interesting *subregular* classes of stringsets



(McNaughton and Papert 1971, Rogers et al. 2010, 2013, Heinz et al. 2011)

LTT	Locally Threshold Testable	TSL	Tier-based Strictly Local
LT	Locally Testable	PT	Piecewise Testable
SL	Strictly Local	SP	Strictly Piecewise

Phonotactics - Knowledge of word well-formedness
Samala Version

ʃtojonowonowaʃ

stojonowonowaʃ

stojonowonowas

ʃtojonowonowas

pisotonosikiwat

pisotonoʃikiwat

asanisotonosikiwasi

aʃanipisotonoʃikiwasi

Phonotactics - Knowledge of word well-formedness
Samala Version

possible Samala words	impossible Samala words
ʃtojonowonowaʃ	stojonowonowaʃ
stojonowonowas	ʃtojonowonowas
pisotonosikiwat	pisotonoʃikiwat
asanisotonoskiwasi	aʃanipisotonoʃikiwasi

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

(Applegate 1972)

Phonotactics - Knowledge of word well-formedness
Samala Version

possible Samala words	impossible Samala words
ʃtojonowonowaʃ	stojonowonowaʃ
stojonowonowas	ʃtojonowonowas
pisotonosikiwat	pisotonoʃikiwat
asanisotonoskiwasi	aʃanipisotonoʃikiwasi

1. Question: How do Samala speakers know which of these words belong to different columns?
2. By the way, *ʃtoyonowonowaʃ* means ‘it stood upright’

(Applegate 1972)

Phonotactics - Knowledge of word well-formedness Language X

possible words of Language X	impossible words of Language X
ʃotkoʃ	sotkoʃ
ʃoʃkoʃ	ʃotkos
ʃosokoʃ	ʃoʃkos
soʃokos	soskoʃ
sokosos	
pitkol	
pisol	
piʃol	

Phonotactics - Knowledge of word well-formedness
Language X

possible words of Language X	impossible words of Language X
ʃotkoʃ	sotkoʃ
ʃoʃkoʃ	ʃotkos
ʃosokoʃ	ʃoʃkos
soʃokos	soskoʃ
sokosos	
pitkol	
pisol	
piʃol	

Sibilant sounds which begin and end words must agree (but not ones word medially).

Phonotactics - Knowledge of word well-formedness Language Y

possible words of Language Y	impossible words of Language Y
ʃotkoʃ	ʃoʃkoʃ
sotkoʃ	ʃoskoʃ
ʃotkos	soʃkos
pitkol	ʃoʃkos
soʃkostoʃ	soskoʃ
	soksos
	piskol
	piʃkol

Phonotactics - Knowledge of word well-formedness
Language Y

possible words of Language Y	impossible words of Language Y
ʃotkoʃ	ʃoʃkoʃ
sotkoʃ	ʃoskoʃ
ʃotkos	soʃkos
pitkol	ʃoʃkos
soʃkostoʃ	soskoʃ
	soksos
	piskol
	piʃkol

Words must have an *even number* of sibilant sounds.

Typology

Attested Phonotactic Patterns

1. Words don't begin with mgl. (English)
2. Words don't contain both ʃ and s. (Samala)

Unattested Phonotactic Patterns

1. Words don't begin and end with disagreeing sibilants. (Language X = First/Last Harmony)
2. Words don't contain an even number of sibilants. (Language Y = *ODD-Sibilants)

What's the explanation?

Optimality Theory

1. Constraints like *#mgl and *[+strident,α anterior]…[+strident,$-\alpha$ anterior] are part of CON.
2. Constraints like *ODD-Sibilants or *#[+strident,α anterior]…[+strident,$-\alpha$ anterior]# are not.

What's the explanation?

Phonetically-based Phonology (Hayes, Kirchner, Steriade 2004)

1. There are perceptual and/or articulatory reasons for constraints like *#mgl and *[+strident,α anterior]...[+strident,$-\alpha$ anterior].
2. There are no such reasons for constraints like *ODD-Sibilants or #[+strident,α anterior]...[+strident,$-\alpha$ anterior]# .

What's the explanation?

Phonetically-based Phonology (Hayes, Kirchner, Steriade 2004)

1. There are perceptual and/or articulatory reasons for constraints like *#mgl and *[+strident,α anterior]…[+strident,$-\alpha$ anterior].
2. There are no such reasons for constraints like *ODD-Sibilants or #[+strident,α anterior]…[+strident,$-\alpha$ anterior]# .

What are those reasons?

First/Last Harmony

1. Long-distance assimilation is well-attested (Hansson 2001, Rose & Walker 2004)
2. Word edges in phonology are privileged positions (Beckman 1997 Fougeron & Keating 1997, Endress, Nespor & Mehler 2009).

Question

What theory of perception or articulation prevents there from being harmony only in privileged positions?

First/Last Harmony

Are the memory requirements greater?

Given the pattern templates, the answer seems to be no.

	[s]	[ʃ]
[s]	✓	✗
[ʃ]	✗	✓

[. . . __. . . __. . .]

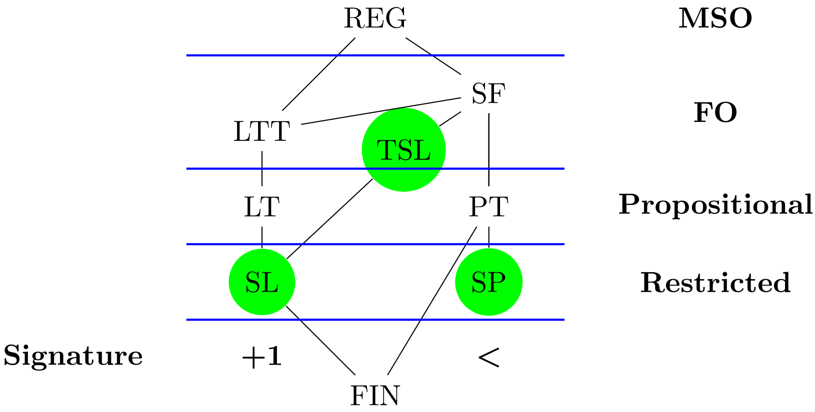
	[s]	[ʃ]
[s]	✓	✗
[ʃ]	✗	✓

[#__. . . __#]

*ODD-Sibilants

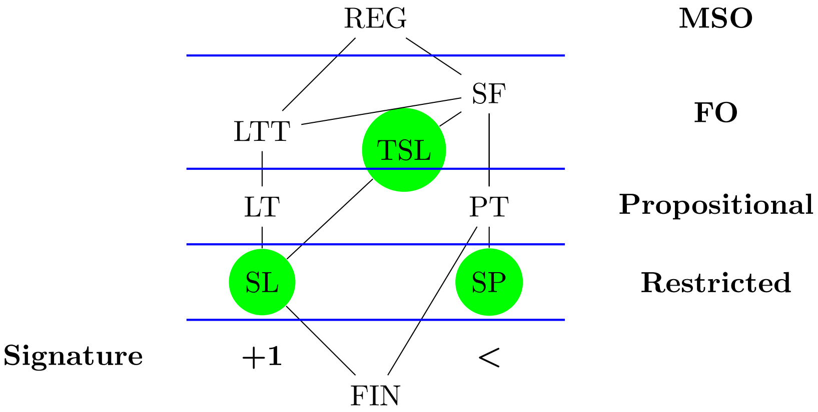
- It's plausible to me at least that perception or articulation should be able to explain the absence of counting mod n patterns in phonology, but I haven't seen any explicit connection.
- Whatever it is, it *should* connect to the computational properties discussed here.

A computational explanation



1. Constraints like *#mgl are Strictly Local.
2. Constraints like *[+strident,α anterior]. . . [+strident,$-\alpha$ anterior] are Strictly Piecewise.

A computational explanation



1. Constraints like First Last Harmony are Locally Testable.
2. Constraints like *ODD-Sibilants are Counting (properly regular).

Other characterizations of the same classes

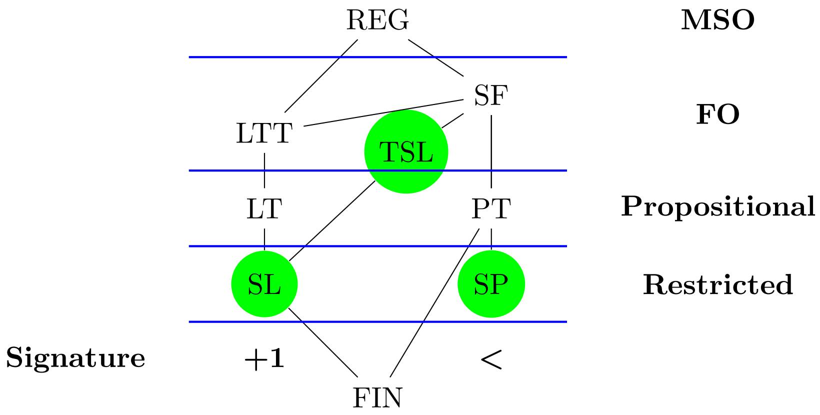
1. Logical characterizations (to be shown)
2. Language-theoretic characterizations
   (independent of any grammar)
3. Element-based grammatical characterization
4. Automata-theoretic characterizations
5. Algebraic characterizations

Other characterizations of the same classes

Engelfriet and Hoogeboom (2001, p.216)

> *It is always a pleasant surprise when two formalisms, introduced with different motivations, turn out to be equally powerful, as this indicates that the underlying concept is a natural one. Additionally, this means that notions and tools from one formalism can be made use of within the other, leading to a better understanding of the formalisms under consideration.*

Logical Signatures



Logical Signatures

The Local Branch (+1)

- (+1) means "successor"
- Literals refer to substrings (contiguous sequences of sounds)

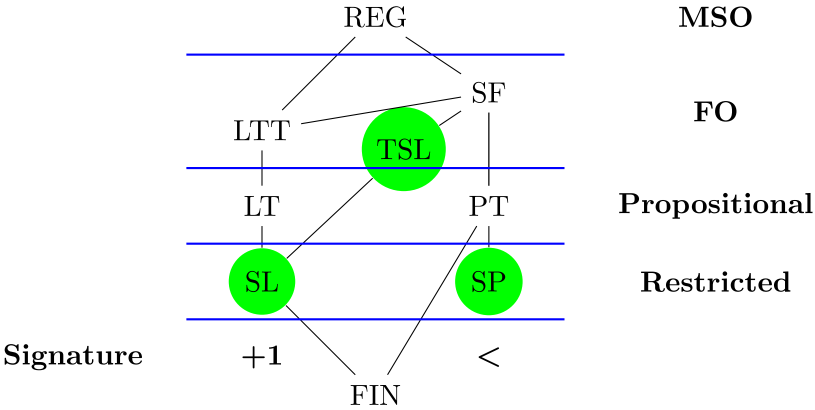
ex. #mgl, VV, …

The Piecewise Branch

- (<) means "precedes"
- Literals refer to subsequences (potentially *dis*contiguous sequences of sounds)

ex. s…s, ʃ…ʃ, a…b …c…

SL and SP: Restricted Logic



SL and SP: Restricted Logic

Finitely many conjunctions of negative literals define stringsets.

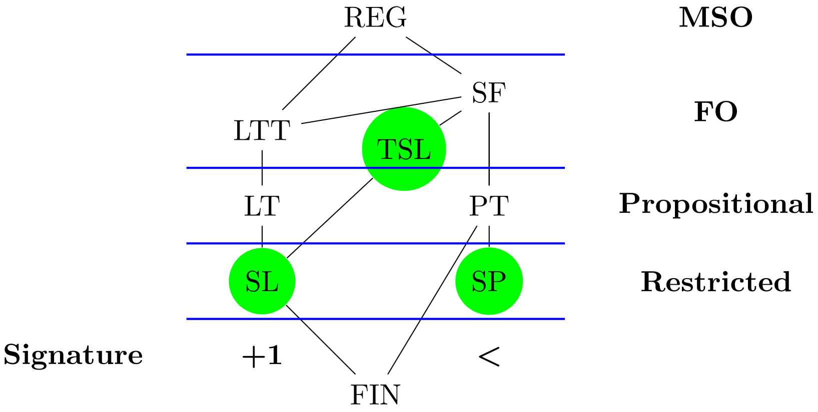
Strictly Local (+1)

example $\neg\#mgl \wedge \neg\#pt \wedge \ldots$
Don't have $\#mgl$ and don't have $\#pt$, ...

Strictly Piecewise (<)

example $\neg\text{s}\ldots\text{ʃ} \wedge \neg\text{ʃ}\ldots\text{s} \wedge \ldots$
Don't have $s\ldots ʃ$ and don't have $ʃ\ldots s$, ...

LT and PT: Propositional Logic



LT and PT: Propositional Logic

Well-formed statements of propositional logic with the literals define stringsets.

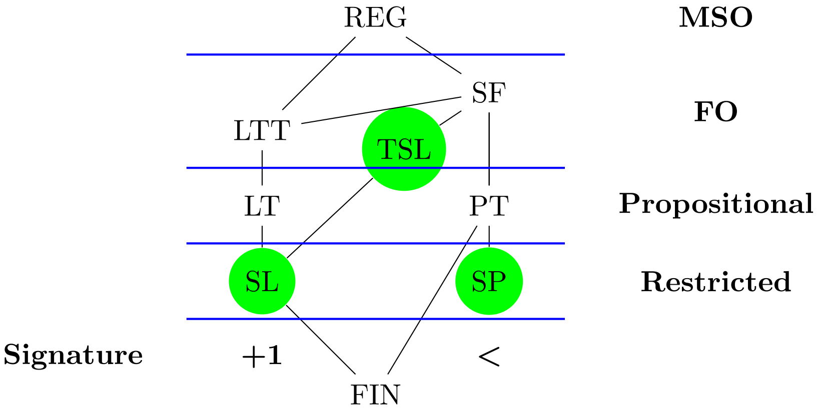
Locally Testable (+1)

example $(\#\mathrm{s} \rightarrow \mathrm{s}\#) \wedge (\#\int \rightarrow \int\#)$
First/Last Harmony

Piecewise Testable (<)

example $\mathrm{s}\ldots\mathrm{s} \rightarrow \int\ldots\int$
If a word has a $s\ldots s$ subsequence, it must also have $\int\ldots\int$ subsequence.

LTT and NonCounting: First Order Logic



LTT and NonCounting: First Order Logic

Well-formed statements of first-order logic with the literals define stringsets. (First order is propositional logic with $\forall, \exists$ quantification over individuals.)

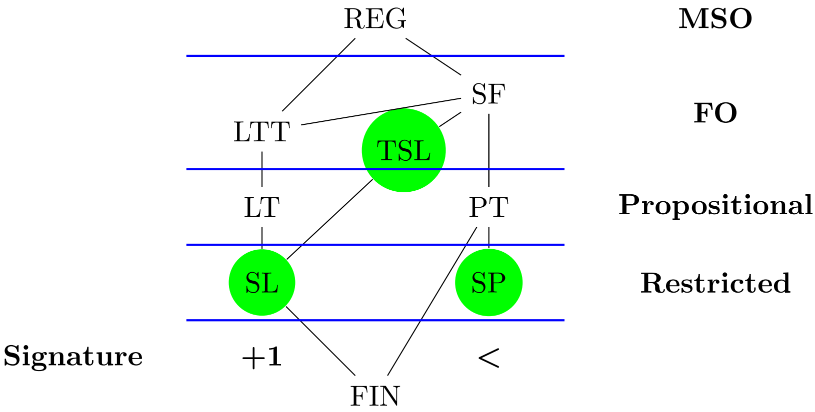
Locally Threshold Testable (+1)

example $\exists(x, y, z)[p(x) \land p(y) \land p(z) \land x \neq y \neq z]$
Words must have three [p]s.

Noncounting (<)

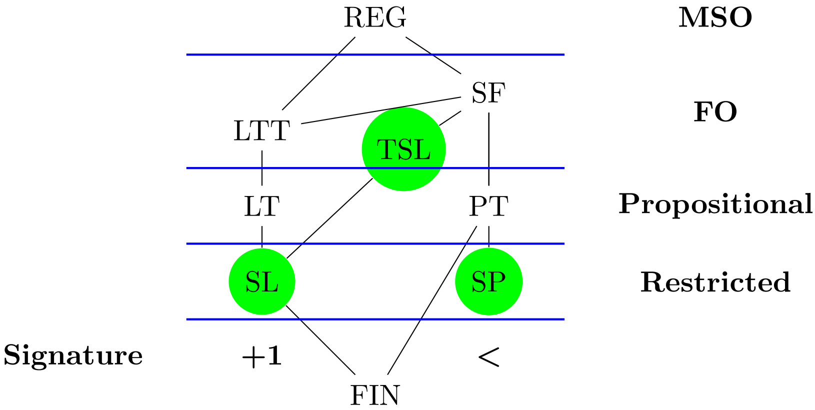
example $(\forall x)\big[s(x) \rightarrow (\exists y)[z(y) \land y < x]\big]$
If a word has [s] then the [s] must be preceded somewhere by a [z].

LTT and Noncounting



"Successor" is first-order definable from "precedence" but not vice versa, which is why Noncounting properly includes LTT.

Regular: Monadic Second Order Logic



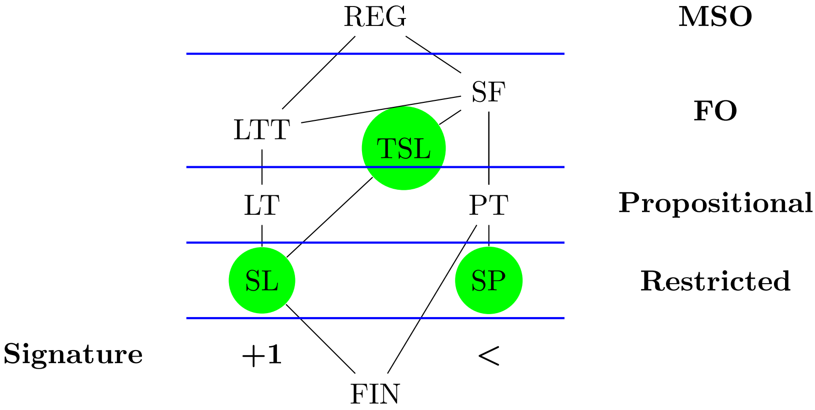
Regular: Monadic Second Order Logic

Well-formed statements of monadic second-order logic with literals from either signature (+1) or (<) define stringsets. (Monadic Second Order is propositional logic with $\forall, \exists$ quantification over *sets* of individuals.)

Regular, either (+1) or (<)

ex. Words must have an even number of sibilants.

Tier-based Strictly Local: Ignoring inconsequential events



Tier-based Strictly Local: Ignoring inconsequential events

Finitely many conjunctions of negative literals *over tiers* define stringsets.

Example

Ignoring nonsibilants

tosopiwaʃonikasan

↓

sʃs

Typology of segmental phonotactic patterns

Phonotactic Patterns derived from

- Constraints on consecutive sequences of sounds are SL
- Long-distance consonantal harmony are both SP and TSL
- Long-distance consonantal disharmony are TSL but not SP
- Vowel harmony without neutral vowels are both SP and TSL
- Vowel harmony with opaque vowels are TSL but not SP
- Vowel harmony with transparent vowels are SP and they are TSL *only* if transparent vowels are off the tier

Heinz 2007, 2010, Rogers et al. 2010, Heinz et al. 2011

Typology of (dominant) Stress Patterns

Of the 109 distinct stress patterns studied in Heinz 2009:

- 9 are SL_2.
- 44 are SL_3.
- 24 are SL_4.
- 3 are SL_5. (Asheninca, Bhojpuri, Hindi (Fairbanks))
- 1 is SL_6. (Icua Tupi)
- 28 are not SL_k for any k. (E.g. unbounded patterns)
- 26 of these are either SP+LT or SL+PT.
- 2 are counting (Cairene Arabic and Creek)

Edlefsen et al. 2009, Graf 2010, Rogers et al. 2012, Heinz to appear, Wibel et al. in prep

Learnability

1. SL_k, SP_k, and $\mathrm{TSL}_{T,k}$ are provably identifiable in the limit from positive data by incremental, set-driven, polytime learning algorithms.

   Garcia et al. 1991, Heinz 2007, 2010, Rogers et al. 2010

   Heinz et al. 2011, Heinz et al. 2012

   - k (and T) must be known a priori.
   - k appears to be small for phonology (perhaps ≤ 5).

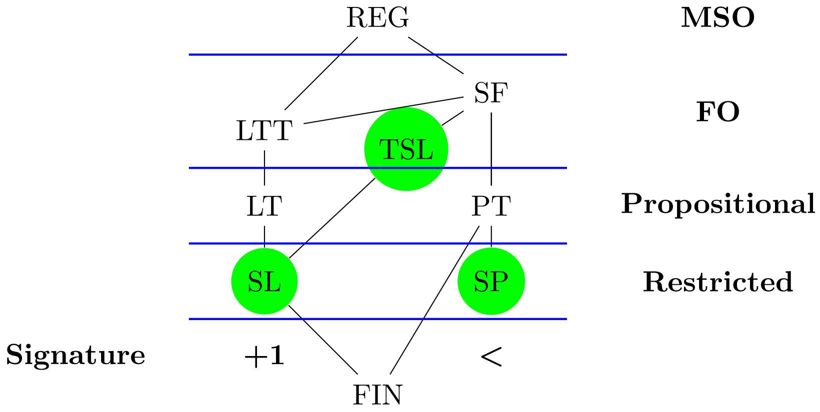
2. Stochastic versions of these algorithms exist which learn probability distributions over stringsets, as well as algorithms incorporating phonological features.

   Jurafsky and Martin 2008, Hayes and Wilson 2008

   Albright 2009, Heinz and Rogers 2010, Heinz and Koirala 2010

A learning explanation

If people generalize from their phonological experience in the ways suggested by these learning procedures then they can only ever learn SL, SP, or TSL patterns.



Psycholinguistic Evidence

Artificial language learning experiments (Lai 2012, 2014)

- Two conditions with the same task at test: forced choice between words
- Sibilant-Harmony condition: familiarized with words obeying the Sibilant-Harmony pattern
- First-Last condition: familiarized with words obeying the First-Last pattern
- Results show subjects in the Sibilant-Harmony internalized the generalization but NOT subjects in the First-Last condition.

What about mappings?

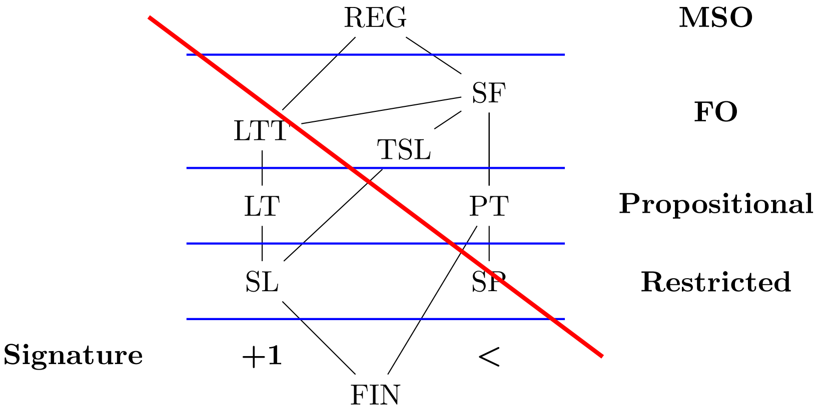
Word-final obstruent devoicing

$$\text{[-sonorant]} \longrightarrow \text{[-voice]} \;/\; \_\#$$

$$*[+\textsc{voice},\text{-}\textsc{sonorant}]\#,\ \textsc{Max}\text{-C} >> \text{ID}(\textsc{voice})$$

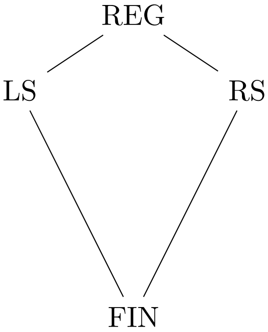
(rat, rat)	(sap, sap)
(rad, rat)	(sab, sap)
. . .	(sag, sat)
(flugenrat, flugenrat)	. . .
(flugenrad, flugenrat)	
. . .	

Regular sets ≠ Regular relations



There are no similar subregular hierarchies for relations (yet)

Subregular Mappings



LS	Left Subsequential	RS	Right-Subsequential

Survey of results for segmental phonology

1. All the iterative vowel harmony patterns described by Nevins (2010) are left or right subsequential (Gainor et al. 2012, Heinz and Lai 2013).
2. All the synchronically attested metathesis patterns, including long-distance ones, in Beth Hume's NSF-funded metathesis database, are left or right subsequential (Chandlee et al. 2012).
3. The typology of partial reduplication patterns in Riggle (2006) are left or right subsequential (Chandlee and Heinz 2012).
4. The long-distance consonantal dissimilation patterns in Suzuki (1998) and Bennett (2013) are left or right subsequential (Payne 2014).
5. The long-distance consonantal harmony patterns in Hansson (2001) are left or right subsequential (Luo, 2014 MS) except for Sanskrit n-retroflexion (Schein and Steriade 1985, Graf 2010).

What is NOT subsequential?

1. The logically possible 'Sour Grapes' vowel harmony pattern (Heinz and Lai 2013).
2. Some alleged long-distance metathesis cases (all diachronic) (Chandlee and Heinz 2012)
3. The common tonal process known as Unbounded Tone Plateauing (UTP), which provides computational evidence for Hyman's hypothesis that "tone is different" from segmental phonology (Jardine 2013).
4. The vowel harmony pattern in Yaka (it's like UTP and as far as we know, unique).

Building the Encyclopedia of Categories (Chandlee 2014)

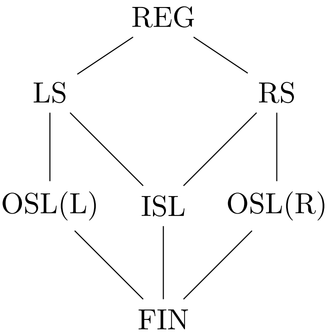
Chandlee 2014 defines Input Strictly Local and Output Strictly Local mappings by synthesizing concepts from

1. Strictly Local stringsets and
2. subsequential functions.

She provides:

1. language-theoretic characterizations and
2. automata-theoretic characterizations.

Building the Encyclopedia of Categories (Chandlee 2014)



LS	Left Subsequential
RS	Right-Subsequential
OSL(L)	Output Strictly Local (Left)
ISL	Input Strictly Local
OSL(R)	Output Strictly Local (Right)

Examining the Encyclopedia of Types (Chandlee 2014)

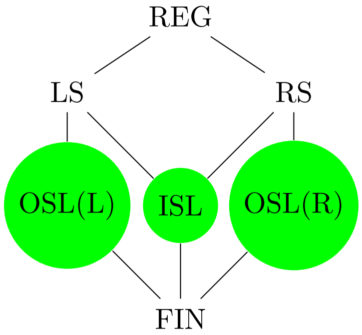
1. Mappings describable with SPE-style rules
   A $\longrightarrow$ B / C __D , where all strings matching CAD are bounded by length k are ISL functions.
   - includes locally-triggered epenthesis, deletion, substitution, and metathesis.
2. Shows that progressive spreading patterns are OSL(L) and regressive ones are OSL(R).
3. Shows that ∼95% of the over ∼5500 processes from over 500 languages in P-base (Mielke 2008) are OSL or OSL.
4. Shows that many word-formation processes are ISL or OSL.
   - includes prefixation, suffixation, infixation, many cases of partial reduplication.
5. Proves that ISL patterns are identifiable from positive data by a learning algorithm which requires significantly fewer computations and requires less data to converge than OSTIA (Oncina et al. 1993).

What's not ISL nor OSL?

Chandlee 2014

1. Long-distance processes like consonant harmony and disharmony...
2. Certain metathesis (displacement) patterns (only attested diachronically).
3. Phonologically bizarre (but subsequential) phonological processes (like sibilant harmony triggered in words by an even number of sibilants).

The typology of phonological mappings



- ISL and OSL mappings approximate the typology of *local* phonological processes reasonably well.

Some Remaining Questions

1. Can we find psycholinguistic evidence that the ISL/OSL boundaries are psychologically real?
2. How can we characterize the Input and Output Strictly Piecewise and Tier-based Strictly Local classes?
3. How can we build a richer Encyclopedia of Categories by considering alternative models of words (so the signatures describe words with features, autosegmental structures, etc.)
4. Phonological generalizations interact. What is the right way to model this interaction (intersection, composition, optimization)?
5. …

Summary

These computational characterizations:

1. Provide an "Encylcopedia of Categories" which can be compared with existing surveys of phonological phenomenon ("The Encyclopedia of Types")
2. Lead to hypotheses about which of the logically possible phonotactic patterns (markedness constraints) and phonological mappings are humanly possible (abstract phonological universals)
3. Can lead to new inductive principles useful for developing learning algorithms (which then becomes the explanans for the universals)

Thanks for listening!