

# Theory Neutral Representations of Stress Patterns

Jeffrey Heinz  
`heinz@udel.edu`

University of Delaware

University of Connecticut  
April 30, 2010

## Goals

1. Generative grammarians characterize stress patterns as infinite sets.
2. Theoretical computer science provides a universal theory of sets.
3. Stress patterns are regular sets, describable with finite-state automata (FSA).
4. The FSA representation is not a theory of stress because it is far too unrestrictive.
5. Formal language theory provides a universal measure which can be fruitfully applied to classify stress patterns

# Outline

Stress patterns as infinite sets

Regular sets

Not a theory of stress

Why we should care

Learning

## Treatment of stress patterns in the typology

The focus is on the dominant stress pattern of the language in terms of syllables.

Typically the following are noted but are not subject to analysis:

1. Lexical exceptions
2. Subsidiary patterns
3. The domain of the stress pattern by word type or prosodic level
4. Morphological factors
5. ...

# Generative grammarians characterize stress patterns as infinite sets

$$\left\{ \begin{array}{l} \acute{\sigma} \\ \acute{\sigma} \sigma \\ \acute{\sigma} \sigma \sigma \\ \acute{\sigma} \sigma \sigma \sigma \\ \dots \end{array} \right\}$$

# Generative grammarians characterize stress patterns as infinite sets

$$\left\{ \begin{array}{l} \acute{\sigma} \\ \acute{\sigma} \sigma \\ \acute{\sigma} \sigma \sigma \\ \acute{\sigma} \sigma \sigma \sigma \\ \dots \end{array} \right\}$$

$$\sigma \longrightarrow [+stress] / \# \text{ —}$$

# Generative grammarians characterize stress patterns as infinite sets

$$\left\{ \begin{array}{l} \acute{\sigma} \\ \acute{\sigma} \sigma \\ \acute{\sigma} \sigma \sigma \\ \acute{\sigma} \sigma \sigma \sigma \\ \dots \end{array} \right\}$$

$$\sigma \longrightarrow [+stress] / \# \text{ —}$$

$$\text{ALIGN}(\text{STRESS}, \text{L}) > \text{ALIGN}(\text{STRESS}, \text{R}), \dots$$

## Pintupi (Quantity-Insensitive Binary)

a.	acute acute	páŋa	‘earth’
b.	acute acute acute	t <sup>j</sup> úŋaya	‘many’
c.	acute acute grave acute	máŋawàna	‘through from behind’
d.	acute acute grave acute acute	púlŋkàlat <sup>j</sup> u	‘we (sat) on the hill’
e.	acute acute grave acute grave acute	t <sup>j</sup> ámulìmpat <sup>j</sup> ùŋku	‘our relation’
f.	acute acute grave acute grave acute acute	t <sup>j</sup> ìlirìŋulàmpat <sup>j</sup> u	‘the fire for our benefit flared’
g.	acute acute grave acute grave acute grave acute	kúran <sup>j</sup> ùlulìmpat <sup>j</sup> ùŋa	‘the first one who is our relation’
h.	acute acute grave acute grave acute grave acute acute	yúma.ŋkamarat <sup>j</sup> ùŋaka	‘because of mother-in-law’
	...		

- Secondary stress falls on nonfinal odd syllables (counting from left)
- Primary stress falls on the initial syllable

Hansen and Hansen (1969:163), Hayes (1995:62)



## Example 2: Latin (Quantity-Sensitive Bounded Single)

- If the penult is heavy (CVC, CV:), stress falls on the penult.  
Otherwise stress falls on the antepenult
- Initial stress in disyllables

L H <b>Á</b> H	L L H <b>Í</b> L L	<b>Á</b> H
H L <b>Á</b> H	<b>Í</b> L H	<b>Í</b> H
L L <b>Á</b> H	L <b>Á</b> L H	<b>Í</b> L
...	...	...

Mester (1992), Jacobs (1989), Hayes (1995:50)

## Example 3: Kwakwala (Quantity-Sensitive Unbounded Single)

Leftmost Heavy Otherwise Rightmost

- Stress the heavy syllable closest to the left edge. If there is no heavy syllable, stress the rightmost syllable.

<b>Ĥ</b> H H	L <b>Ĭ</b>
L L <b>Ĥ</b> L L	L L <b>Ĭ</b>
L L L <b>Ĥ</b>	L L L <b>Ĭ</b>
...	...

Zec (1994), Walker (2000:21)

## Kinds of Attested Stress Patterns

- Combining Bailey's 1995 survey and Gordon's 2002 survey yields a survey of 405 languages (423 descriptions and 109 distinct patterns).
  - Quantity-Insensitive: Single, Dual, Binary, Ternary
  - Quantity-Sensitive Bounded: Single, Dual, Binary, Ternary, Multiple
  - Quantity-Sensitive Unbounded: Single, Binary, Multiple
- Secondary stress patterns for QS patterns were collected from secondary and primary sources with help of Stephen Tran and Rachel Schwartz (UCLA Ling undergrads in 2006-2007).

# Stress Patterns Not Found in the World's Languages

- 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
- ...

## The Question

What kind of infinite sets are they?

- Despite the extensive variation, stress patterns are not arbitrary.
- Understanding the limits of this variation is, and continues to be, one of the goals of generative phonology.

(Chomsky and Halle 1968, Liberman 1975, Hyman 1977, Prince 1980, Hayes 1981, Prince 1983, Halle and Vergnaud 1987, Idsardi 1992, Prince 1992, Bailey 1995, Hayes 1995, Hammond 1999, Elenbaas and Kager 1999, Walker 2000, Hyde 2002, Gordon 2002, Baković 2004, R. Goedemans 2005, Hyde 2007, H.G. van der Hulst 2009, R. Goedemans, H.G. van der Hulst & E. van Zanten in press, ...)

# Outline

Stress patterns as infinite sets

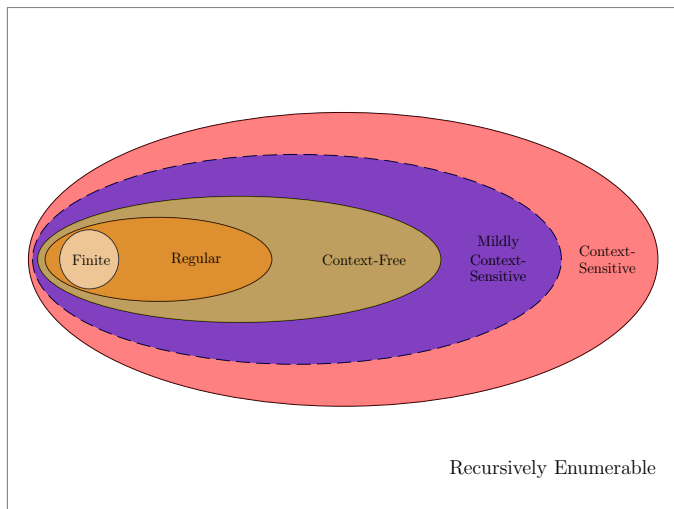
**Regular sets**

Not a theory of stress

Why we should care

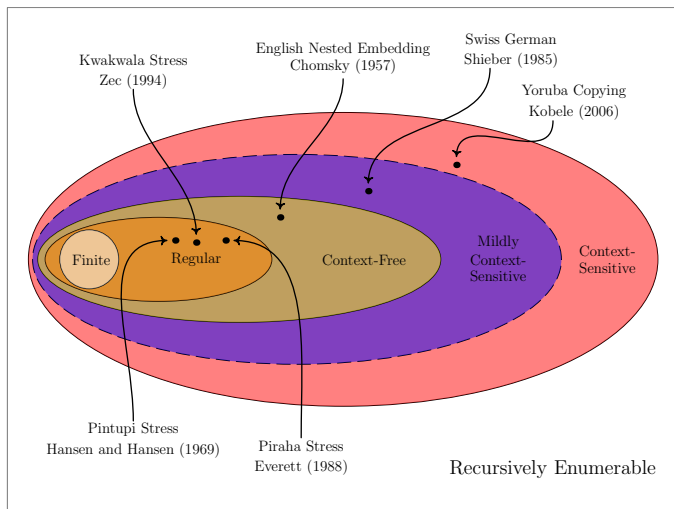
Learning

# The Chomsky Hierarchy



1. Formal language theory is a universal theory of infinite sets.
2. Each region has multiple, independent, converging characterizations.

# The Chomsky Hierarchy



1. Formal language theory is a universal theory of infinite sets.
2. Each region has multiple, independent, converging characterizations.



## Stress patterns are regular sets

Regular sets have many independently motivated, converging characterizations (Harrison 1978, Kracht 2003).

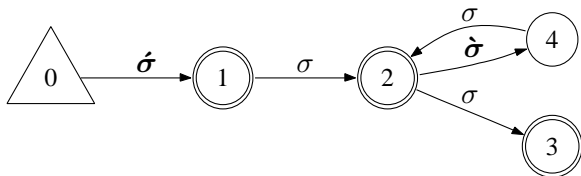
1. Right-branching rewrite rules
2. Formula in Monadic Second Order Logic
3. Finite State Automata (FSAs)
4. Regular expressions
5. ...

## Stress patterns are regular sets

Regular sets can be characterized with Finite State Acceptors (FSAs). These

1. accept or reject words. So it meets the minimum requirement for a phonotactic grammar– a device that at least answers Yes or No when asked if some logically possible word is possible.  
(Chomsky and Halle 1968, Halle 1978)
2. can be computed from SPE-style grammars  
(Johnson 1972, Kaplan and Kay 1994)
3. can be computed from OT grammars  
(Frank and Satta 1998, Karttunen 2000, Riggle 2004)
4. are well-defined and can be manipulated  
(Hopcroft et. al. 2001).

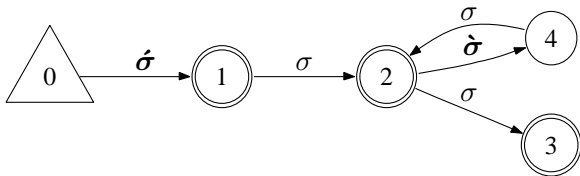
# Example 1: Pintupi (Quantity-insensitive Binary)



Accepts	Rejects
$\acute{\sigma}$	
$\acute{\sigma} \sigma$	
$\acute{\sigma} \sigma \sigma$	
$\acute{\sigma} \sigma \grave{\sigma} \sigma$	
$\acute{\sigma} \sigma \grave{\sigma} \sigma \sigma$	
...	

- Every word the machine accepts obeys the rule.
- Every word the machine rejects violates it.

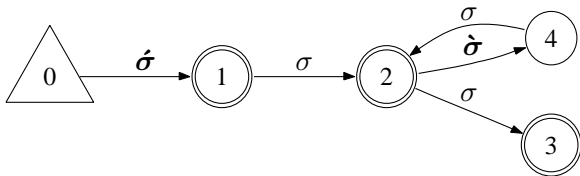
# Example 1: Pintupi (Quantity-insensitive Binary)



Accepts	Rejects
	$\acute{\sigma} \sigma \grave{\sigma}$
	$\acute{\sigma} \sigma \sigma \grave{\sigma} \sigma$
	$\sigma \sigma \sigma \sigma \acute{\sigma} \sigma$
	...

- Every word the machine accepts obeys the rule.
- Every word the machine rejects violates it.

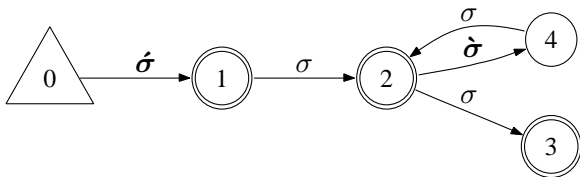
# Example 1: Pintupi (Quantity-insensitive Binary)



Accepts	Rejects
$\acute{\sigma}$	$\acute{\sigma} \sigma \grave{\sigma}$
$\acute{\sigma} \sigma$	$\acute{\sigma} \sigma \sigma \grave{\sigma} \sigma$
$\acute{\sigma} \sigma \sigma$	$\sigma \sigma \sigma \sigma \acute{\sigma} \sigma$
$\acute{\sigma} \sigma \grave{\sigma} \sigma$	...
$\acute{\sigma} \sigma \grave{\sigma} \sigma \sigma$	
...	

- Every word the machine accepts obeys the rule.
- Every word the machine rejects violates it.

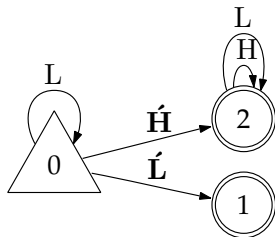
## Example 1: Pintupi (Quantity-insensitive Binary)



- The (different) OT analyses of the Pintupi pattern in Gordon (2002) and Tesar (1998) describe the same infinite set as the acceptor above.

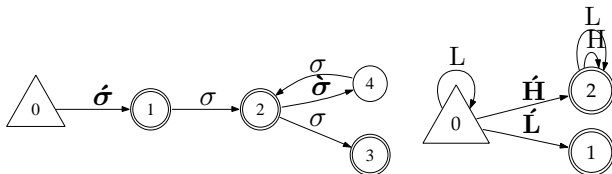
## Example 2: Kwakwala (Quantity-sensitive Unbounded)

Leftmost Heavy Otherwise Rightmost



- The (different) OT analyses of the Kwakwala pattern in Walker (2000) and Baković (2004) describe the same infinite set as the acceptor above.

## Local Summary



- **All** known stress patterns can be represented with finite state acceptors (even Pirahã).
- These grammars recognize an infinite number of legal words, just like the generative grammars of earlier researchers.



# Outline

Stress patterns as infinite sets

Regular sets

**Not a theory of stress**

Why we should care

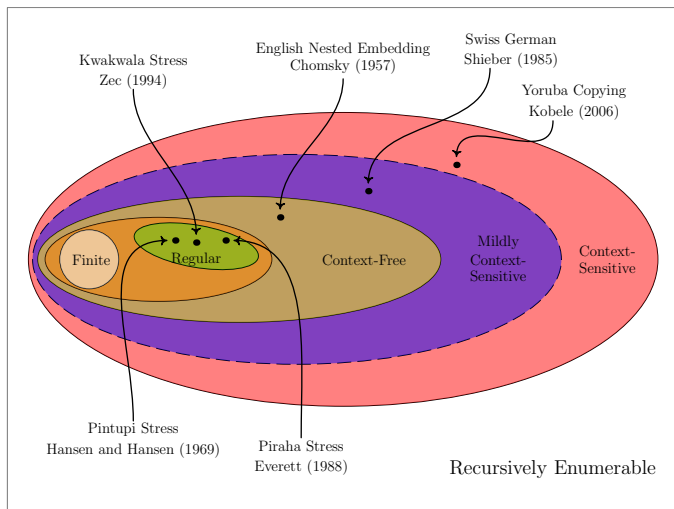
Learning

# A theory that all stress patterns are regular is weak

Many phonologically bizarre patterns can be written with finite state acceptors.

1.	2.	3.
<i>ó</i>	<i>ó</i>	<i>ó</i>
<i>σ ó</i>	<i>ó σ</i>	<i>ó σ</i>
<i>ó σ σ</i>	<i>ó σ ò</i>	<i>ó σ σ</i>
<i>σ σ σ ó</i>	<i>ó σ ò σ</i>	<i>σ σ σ σ</i>
<i>ó σ σ σ σ</i>	<i>ó σ ò σ σ</i>	<i>σ σ σ σ σ</i>
<i>σ σ σ σ σ ó</i>	<i>ó σ ò σ σ σ</i>	...
...	<i>ó σ ò σ σ σ σ</i>	
	...	

# Theories of Stress



1. Establish the limits of variation
2. Seek an explanation for such limits

3. Several factors: articulatory, perceptual, formal, ...

# Outline

Stress patterns as infinite sets

Regular sets

Not a theory of stress

**Why we should care**

Learning

## Some reasons already mentioned

### FSA representations

1. Are practical: simple, supported by many algorithms for word generation, recognition and combination
2. Allows phonologists working in different theories to automatically compare specific predictions by translating theories into FSAs.
3. Helps focus attention on the infinite sets which are the generalizations we aim to describe.

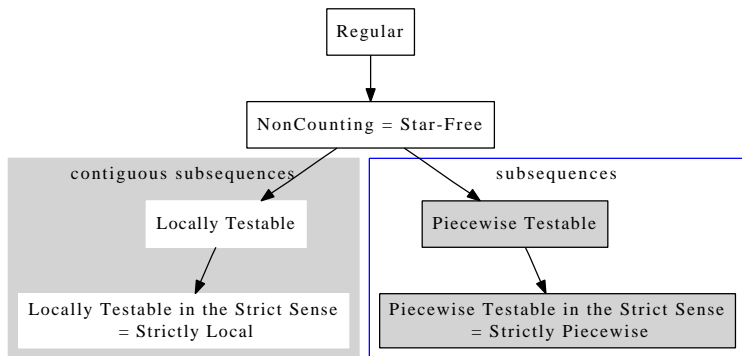
## Some reasons not yet mentioned

1. FSAs are a lingua franca across fields and time.
2. Additional classification exists within regular sets!
3. Interesting properties can be identified with these representations (e.g. to help the learning problem)

## FSAs are a lingua franca across fields and time

1. Other fields are interested in regular sets for different reasons in different domains (theoretical computer science, robotics, psychology, ...)
2. Facilitates communication between phonologists and researchers in other fields
3. Facilitates comparison to the regular sets in other domains
4. FSAs (and other canonical representations of regular sets, more generally) stand the test of time. People will still use them in 100+ years.

# Classifying regular sets: the Subregular Hierarchies



1. Each class has multiple, independently-motivated characterizations (McNaughton and Papert 1971, Simon 1975, Rogers et. al to appear, Rogers and Pullum 2009, Heinz to appear)



# Classifying Stress Patterns in the Subregular Hierarchies

Edlefsen et. al 2008:

1. QI binary patterns are Strictly 2-Local.
2. QI ternary patterns are Strictly 3-Local.
3. Antepenultimate patterns are Strictly 4-Local.
4. Içuã Tupi is Strictly 5-Local.
5. QS Bounded patterns are also Strictly Local.
6. QS Unbounded patterns are Noncounting.

Graf 2009:

1. Creek and Cairene Arabic are not Noncounting. (They count modulo 2).
- Also: No stress patterns are in the Piecewise branch.

# Outline

Stress patterns as infinite sets

Regular sets

Not a theory of stress

Why we should care

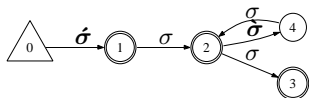
**Learning**

# Understanding the Learning Problem

The central problem in learning is generalization

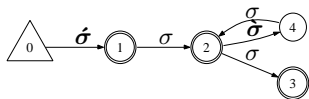
1. Learners generalize beyond their finite experience.
2. This is only possible because their hypothesis space is a priori restricted.
3. What kinds of hypothesis spaces include attested stress patterns and are learnable?

# The Learning Question in Context



- How can this finite state acceptor be learned from a finite list of Pintupi words?
- This question is not easy. There is no simple ‘fix’.

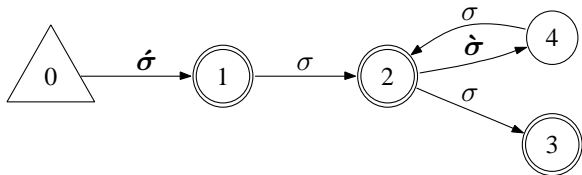
## The Learning Question in Context



- How can this finite state acceptor be learned from a finite list of Pintupi words?
- This question is not easy. There is no simple ‘fix’.

In fact, it is known that no learner can learn the class of regular sets (Gold 1967, Jain et. al 1999)

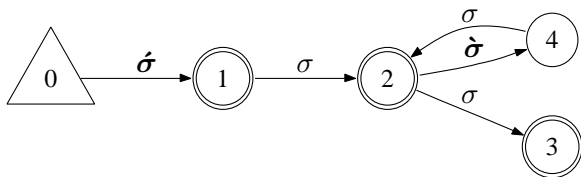
## The Answer to the Learning Question



**Q:** How can this finite state acceptor be learned from the finite list of Pintupi words?

- A:**
- Generalize by writing smaller and smaller descriptions of the observed forms
  - guided by some universal property of the target class...

## The Answer to the Learning Question



**Q:** How can this finite state acceptor be learned from the finite list of Pintupi words?

**A:**

- Generalize by writing smaller and smaller descriptions of the observed forms
- guided by some universal property of the target class... **locality!**

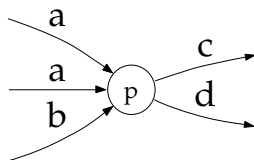
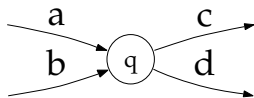
## Neighborhoods (Heinz 2009)

- 107 of the 109 stress patterns have have acceptors whose states which are uniquely identified by their local environment.
- The two exceptions Içuã Tupi and Hindi (per Kelkar).
- I call this environment the *neighborhood*. It is:
  - (1) the set of incoming symbols to the state
  - (2) the set of outgoing symbols to the state
  - (3) whether it is a final state or not
  - (4) whether it is a start state or not



## Example of Neighborhoods

- States  $p$  and  $q$  have the same neighborhood.



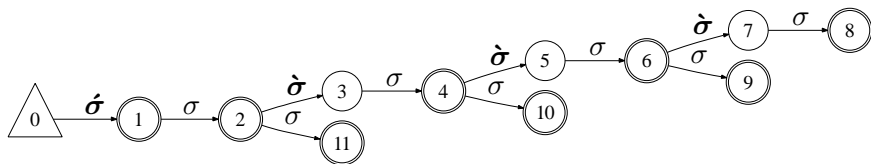
## Neighborhood-distinctness

- A language (regular set) is *neighborhood-distinct* iff there is an acceptor for the language such that each state has its own unique neighborhood.
- 107 of the 109 stress patterns are neighborhood-distinct.
- This makes *neighborhood-distinctness* a (near) universal of attested stress patterns.

# Overview of the Neighborhood Learner

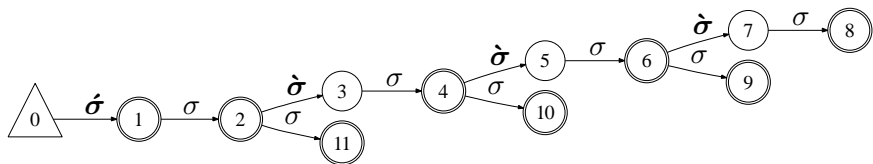
1. Builds a structured representation of the input list of words
2. Generalizes by merging states which are redundant:  
i.e. those that have the same local environment—**the neighborhood**

## The Prefix Tree for Pintupi Stress



- Accepts the words:  $\boxed{\acute{\sigma}}$ ,  $\boxed{\acute{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma}}$ ,  
 $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma}}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma} \sigma}$ ,  
 $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma}}$
- A structured representation of the input.
- It accepts only the forms that have been observed.
- Note that environments are **repeated** in the tree!

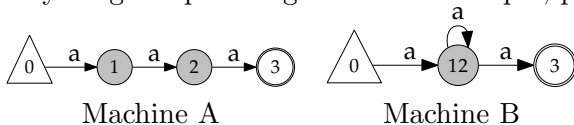
## The Prefix Tree for Pintupi Stress



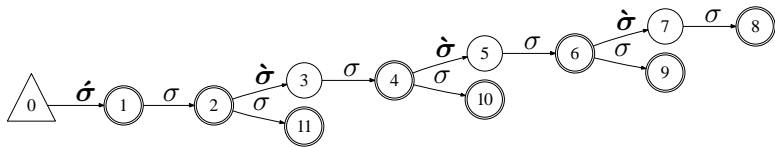
- Accepts the words:  $\boxed{\acute{\sigma}}$ ,  $\boxed{\acute{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma}}$ ,  
 $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma}}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma} \sigma}$ ,  
 $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma}}$
- A structured representation of the input.
- It accepts only the forms that have been observed.
- Note that environments are **repeated** in the tree!

## Generalizing by State-merging

- Eliminate redundant environments by *state-merging*.
- This is a process where two states are identified as equivalent and then *merged* (i.e. combined).
- A key concept behind state merging is that transitions are preserved (Angluin 1982, Muggleton 1990).
- This is one way in which generalizations may occur—because the post-merged machine accepts everything the pre-merged machine accepts, possibly more.



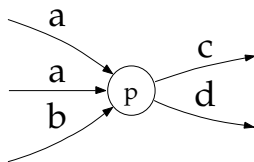
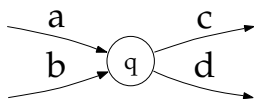
## The Learner's State Merging Criteria



- How does the learner decide whether two states are equivalent in the prefix tree?
- Merge states with the same neighborhood.

## Example of Neighborhoods

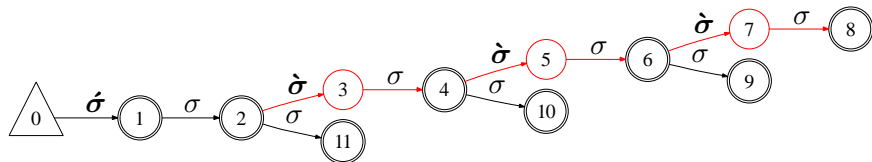
- States  $p$  and  $q$  have the same neighborhood.



- The learner merges states in the prefix tree with the same neighborhood.

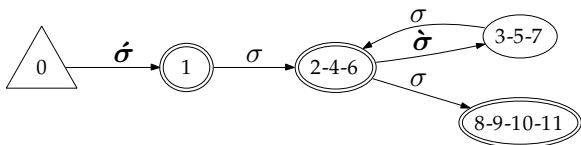


## The Prefix Tree for Pintupi Stress



- States 3, 5, and 7 have the same neighborhood.
- So these states are merged.

## The Result of Merging Same-Neighborhood States



- The machine above accepts
  - $\boxed{\acute{\sigma}}$ ,  $\boxed{\acute{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \sigma}$ ,  $\boxed{\acute{\sigma} \sigma \grave{\sigma} \sigma \grave{\sigma} \sigma}$ ,
  - ...
- The learner has acquired the stress pattern of Pintupi, i.e. it has generalized exactly as desired.
- Each state in the acceptor above has a **distinct neighborhood**.

## Summary of the Forward Neighborhood Learner

- (1) Builds a prefix tree of the observed words.
- (2) Generalize by merging states which have the same neighborhood (local environment).
- (3) The acceptor returned by the algorithm is **neighborhood-distinct**—every state has a distinct neighborhood.

## Summary of Results

- The learner succeeds for 100 of the 109 patterns
  - 37 of the 39 quantity-insensitive patterns
  - 38 of the 44 quantity-sensitive bounded patterns
  - 25 of the 26 quantity-sensitive unbounded patterns

## Why It Works: Neighborhood-distinctness

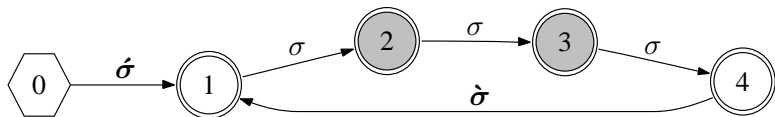
- Every attested stress pattern is neighborhood-distinct except Icuá Tupi and Hindi (per Kelkar) (this can be verified upon inspection).
- Because the learner merges states with the same neighborhood, it learns neighborhood-distinct patterns (except Pirahã, Asheninca, ...).
- Thus, the learner is really taking advantage of this previously unnoticed universal property of these grammars: **neighborhood-distinctness**

## Patterns not Learnable by the Neighborhood 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
- ...

These patterns are **not** neighborhood-distinct.

## Example: Quaternary Stress



- States 2 and 3 have the same neighborhood.
- It is not possible to write any finite state machine for this stress pattern where no two states have the same neighborhood.
- The learner fails because it does not distinguish in some sense “exactly three” from “more than two.”
- **It cannot count past two.**

## Locality in Phonology

- “Consider first the role of counting in grammar. How long may a count run? General considerations of locality, ... suggest that the answer is probably ‘up to two’: a rule may fix on one specified element and examine a structurally adjacent element and no other.” (McCarthy and Prince 1986:1)
- “... the well-established generalization that linguistic rules do not count beyond two ...” (Kenstowicz 1994:597)
- “... it was felt that phonological processes are essentially local and that all cases of nonlocality should derive from universal properties of rule application” (Halle and Vergnaud 1987:ix)
- “Metrical theory forms part of a general research program to define the ways in which phonological rules may apply non-locally by characterizing such rules as local with respect to a particular representation.” (Hayes 1995:34)



## Learnable Unnatural Patterns

- There are stress patterns that can be learned by neighborhood learning which are not considered natural.
  - (1) Leftmost Light otherwise Rightmost.
  - (2) A stress pattern requiring both lapses and clashes.
  - (3) A stress pattern where all syllables have primary stress.
- Do we expect the explanation for these unattested patterns to follow from considerations of locality?

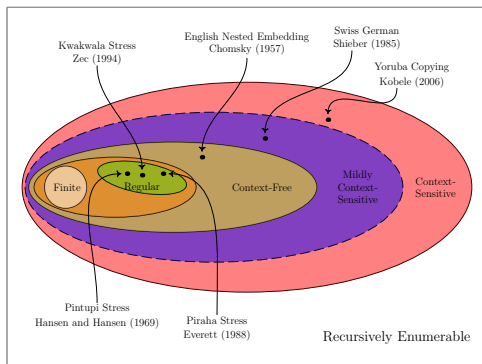
## Learnable Unnatural Patterns

- There are stress patterns that can be learned by neighborhood learning which are not considered natural.
  - (1) Leftmost Light otherwise Rightmost.
  - (2) A stress pattern requiring both lapses and clashes.
  - (3) A stress pattern where all syllables have primary stress.
- Do we expect the explanation for these unattested patterns to follow from considerations of locality?

## Locality is But One Factor in Learning

- This work belongs to a larger research program which is to identify and isolate properties of natural language which are helpful to learning.
- We should ask: What other properties exist
  - which better approximate the class of possible stress systems?
  - and which might assist learning?
- Neighborhood-distinctness is restrictive: most logically possible regular sets do not have this property.

## Summary



1. Stress patterns are regular patterns.
2. Standard representations of regular sets help us focus on the generalizations themselves (the infinite sets), understand differences and similarities between different phonological theories, have longevity, ...
3. Exploring the subregular properties of stress patterns can be a fruitful approach to typology as well as the learning problem.