Goldilocks meets the subset problem:

Evaluating Error Driven Constraint Demotion for OT language acquisition

Joshua Tauberer
University of Pennsylvania

PLC32 – 2/23/08

Questions to Answer

- “UG” defines the initial states of the learner
  - What do different possibilities for UG say about which languages are learnable?
  - “UG” must be only as narrow as we need to account for what languages are learnable
- How do we measure learnability?
  - What is a “successful learner”?
  - The notion of “success” should not be too broad or too narrow

Overview

- Error Driven Constraint Demotion
- Other approaches to OT learning
- Markov model
- What is “Success”?
- Simulations
- Conclusion

EDCD

Error Driven Constraint Demotion

- Tesar and Smolensky (1996, 2000)
- Learner maintains one hypothesis
- Demotes constraints in response to input

- I simulate using a Markov model.

Oral/Nasal Vowel Alternation

- English pattern:

base	output
pæt	pæt
pænt	pæ̃nt
pæ̃t	pæt
pæ̃nt	pæ̃nt

$*V_{ORAL}N \gg *V_{NASAL} \gg$ IDENT-IO(nasal)

Kager (1999), p31

Error Driven Constraint Demotion

hypothesis before: *nas ≫ *voice ≫ *V$_{ORAL}$N ≫ ID(nas)

Learner hears: "pæ̃nz"

Robust Interpretive Parsing: "pæ̃nz"

Run Tableau

Constraint Demotion

*nas: No nasal vowels.

*V$_{ORAL}$N: No oral vowels followed by nasal consonants.

?	*nas	*voice	*V$_{ORAL}$N	ID(nas)
pænz		*	*	?
pæ̃nz	*!	*		?

ID(nas): A manner faithfulness constraint for nasality (for vowels).

hypothesis after: *voice ≫ *V$_{ORAL}$N ≫ { *nas , ID(nas) }

Error Driven Constraint Demotion

hypothesis before: *nas » *voice » *VoralN » ID(nas)

Learner hears: "pæ̃nz"

Robust Interpretive Parsing: "pæ̃nz"

Run Tableau

Constraint Demotion

What do we use as the base form?

?	*nas	*voice	*VoralN	ID(nas)
pænz		*	*	?
pæ̃nz	*!	*		?

Robust Interpretive Parsing:
Use what you heard.

hypothesis after: *voice » *VoralN » { *nas , ID(nas) }

Error Driven Constraint Demotion

hypothesis before: *nas » *voice » *VoralN » ID(nas)

Learner hears: “pæ̃nz”

Robust Interpretive Parsing: “pæ̃nz”

Run Tableau

Constraint Demotion

Child’s Form “Loser”

pæ̃nz	*nas	*voice	*VoralN	ID(nas)
► pænz		*	*	*
pæ̃nz	*!	*		

Adult Form “Winner”

hypothesis after: *voice » *VoralN » { *nas , ID(nas) }

Error Driven Constraint Demotion

hypothesis before: *nas ≫ *voice ≫ *VORALN ≫ ID(nas)

Learner hears: "pæ̃nz"

Robust Interpretive Parsing: "pæ̃nz"

Run Tableau

Constraint Demotion

Child's Form "Loser"

pæ̃nz	*nas	*voice	*VORALN	ID(nas)
▶ pænz		~~*~~	*	*
pæ̃nz	*!	~~*~~		

Mark Cancellation

Highest Uncanceled Loser Mark

hypothesis after: *voice ≫ *VORALN ≫ { *nas , ID(nas) }

Error Driven Constraint Demotion

hypothesis before: *nas ≫ *voice ≫ *VoralN ≫ ID(nas)

Learner hears: "pæ̃nz"

Robust Interpretive Parsing: "pæ̃nz"

Run Tableau

Constraint Demotion

All Uncanceled Winner Marks

pæ̃nz	*nas	*voice	*VoralN	ID(nas)
▶ pænz		~~*~~	*	*
pæ̃nz	*!	~~*~~		

hypothesis after: *voice ≫ *VoralN ≫ { *nas , ID(nas) }

Error Driven Constraint Demotion

hypothesis before: *nas ≫ *voice ≫ *VORALN ≫ ID(nas)

Learner hears: “pæ̃nz”

Robust Interpretive Parsing: “pæ̃nz”

Run Tableau

Constraint Demotion

All Uncanceled Winner Marks

pæ̃nz	*nas	*voice	*VORALN	*nas	ID(nas)
pænz		~~*~~	*!		*
▶ pæ̃nz	*!	~~*~~		*	

hypothesis after: *voice ≫ *VORALN ≫ { *nas , ID(nas) }

'Robust' Interpretive Parsing

<u>Metrical Stress</u>

Underlying Forms: HHLHL

Candidates: · (`·)(`·)

Output Forms: H**H**L**H**L

Underlying forms are transparent from output forms.

<u>Phonological Alternation</u>

Underlying Forms: /pænz/

Candidates: /pæ̃nz/

Output Forms: /pæ̃nz/

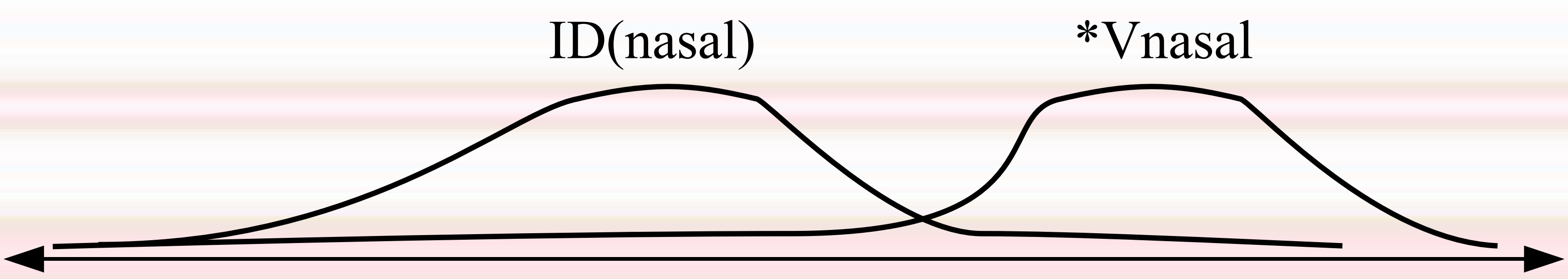
Underlying forms are not transparent.

- Can't demote faithfulness constraints
  - Must bias the system (not the concern here)

Other Approaches to OT Learning

Gradual Learning Algorithm

- Constraints ranked on a continuous rather than discrete scale
- Gradual nudges: robust to erroneous adults
- Assumes the learner sees underlying forms



Boersma (1997), Boersma & Hayes (2001)

Multi-Recursive CD

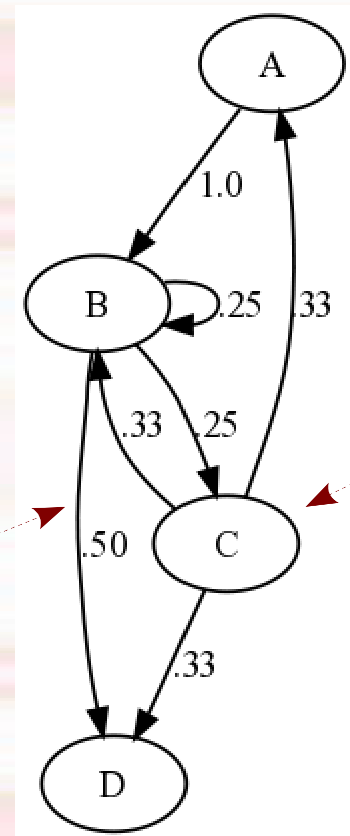
- Learner maintains a *set* of hypotheses.
  - Amenable to biases to address subset problem
- Learner tries every possible parse of each input and expands his set of hypotheses:
  - Exponential growth? Psychologically plausible?

Tesar (1997)

Markov Models
for Language Learning

Markov Models

- Discrete states
- Transition probabilities
- Move to next state does not depend on history



*nas ≫ *voice ≫ *VoralN ≫ ID(nas)

*voice ≫ *VoralN ≫ { *nas , ID(nas)

Probability learner with hypothesis B hears a word that takes him to hypothesis D.

(Depends on language in envir-onment.)

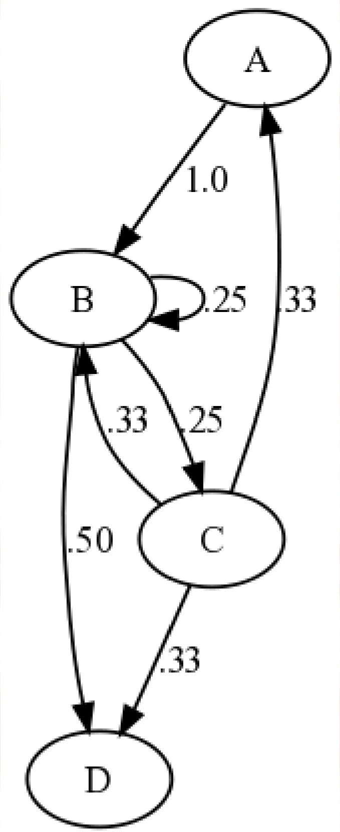
Niyogi and Berwick (1996)

Markov Model Details

- M_{ij} is the transition probability from state i to state j

...to

from...		A	B	C	D
	A	0	1	0	0
from...	B	0	¼	¼	½
	C	⅓	⅓	0	⅓
	D	0	0	0	1



Markov Model Details

- M^t captures the complete behavior of a learner given a distribution of overt adult forms
  - Does not depend on a particular order of input
  - Considers all initial states simultaneously
- Questions this answers:
  - From initial state i, which grammar(s) does the learner converge to?
  - For a target grammar j, where can the learner start so he reaches j?

What is success?

(Goldilocks point 1)

"Success" in Learning

- Four notions of success:
  - ~~Matching the adult grammar exactly~~
  - "Functional" Success
    - Same input-output mapping; different constraint order
  - "Extensional" Success
    - Same set of overt forms; maps to base forms differently
  - "Superset" Success
    - All adult overt forms, and then some

Simulations

Simulation 1

- Six constraints to explain two phenomena
  - Oral/Nasal Vowels in English (as before)
  - Denasalization in Mandar (Kager 1999, p81)
    - /N/ → /t/ before an unvoiced consonant:

/maN-dundu/	mandundu 'to drink'
/maN-tunu/	mattunu 'to burn'

Simulation 1 Constraints

- *nas
- $*V_{ORAL}N$
- ID_V(nas)
- *NC̥: No nasal consonants before an unvoiced consonant
- ID(voice): Voice faithfulness
- ID_C(nasality): Nasality faithfulness for consonants

Simulation 1 Methods

- 6! = 720 fully stratified grammars
- For each grammar, the input-output mapping was computed automatically

base	output
pæt	?
pæ̃t	?
pænt	?
pæ̃nt	?
pæd	?
pæ̃d	?
pænd	?
pæ̃nd	?

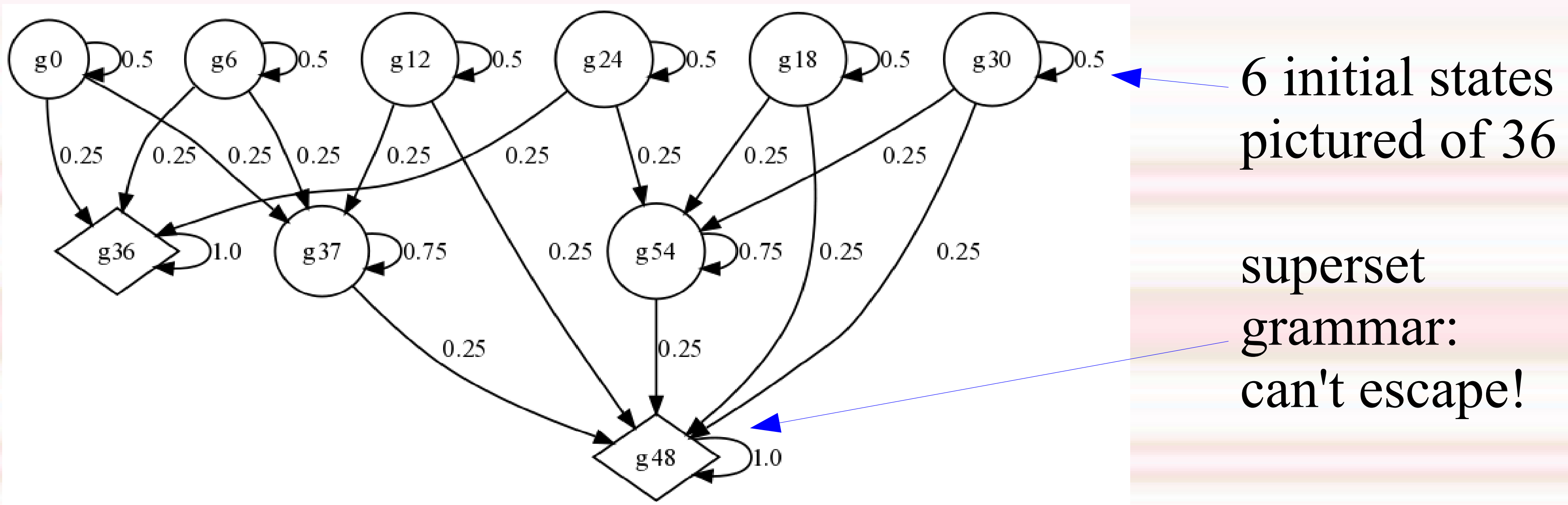
- "Richness of the Base" to devise the lexicon
- All lexical entries assumed occur with equal frequency

Simulation 1 Methods

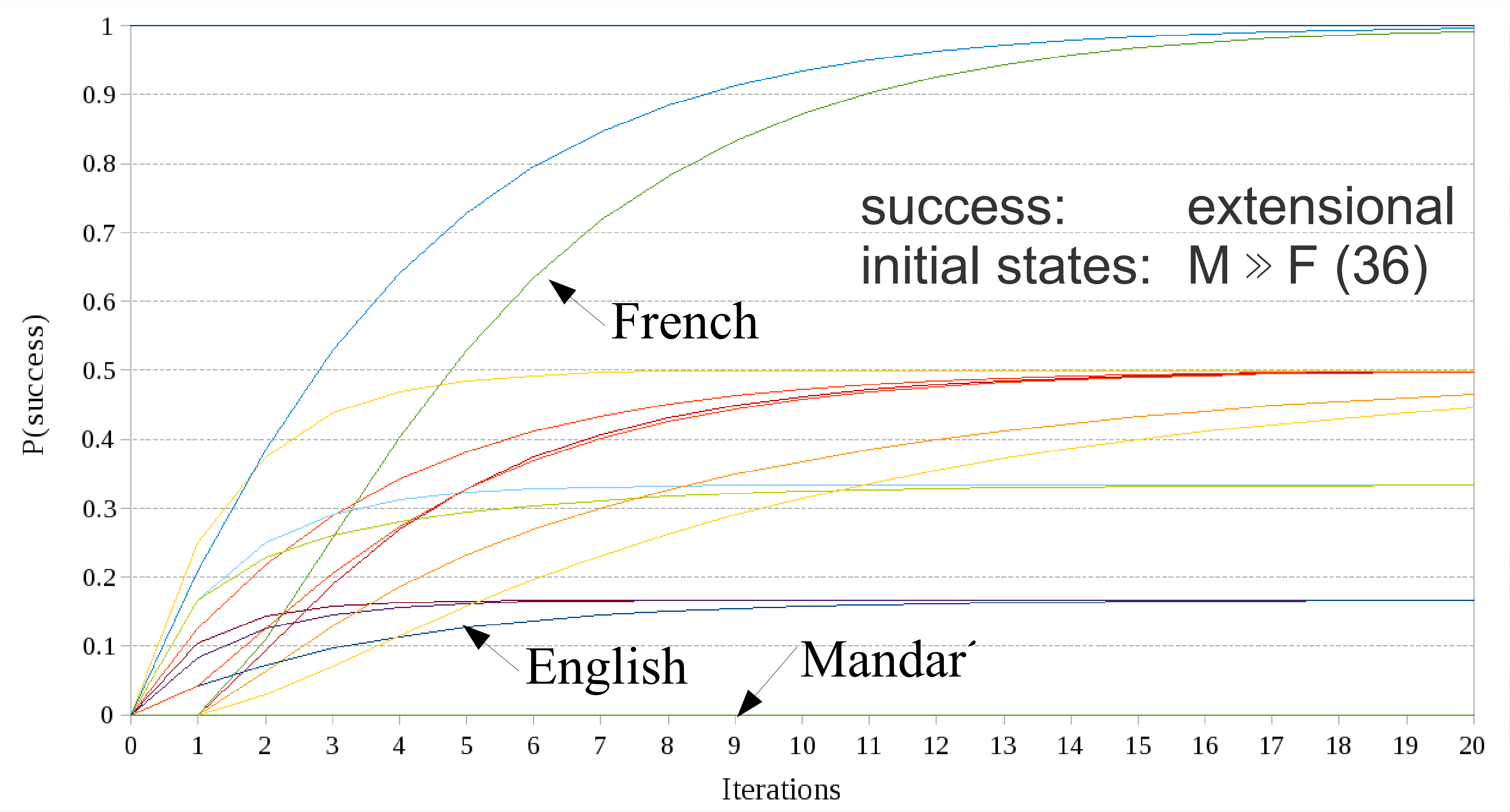
- The adult languages considered:
  - The 16 out of 720 grammars that were ‘functionally’ distinct
- Several sets of initial states were used
  - “UG”

State Transition Diagram

- English target language (sets transition probs.)
- Initial states considered: mark'dness » faith'ness
- Convergence is fast! But to what grammar?



Rate of Convergence



Simulation 1 Results

"Extensional" Success

UG	Lx. Learnable	P(English Learned)
all (720)	1/16	14%
M » F (36)	3/16	17%
{M} » {F}	7/16	0%
{ all }	1/16	0%

But English *is* learnable, so either:

Extensional Success is too strong, or

EDCD does not work

Simulation 1 Results

“Superset” Success

UG	Lx. Learnable	P(English Learned)	P(Mandar´ Learned)
all (720)	8/16	100%	46%
M » F (36)	9/16	100%	67%
{M} » {F}	16/16	100%	100%
{ all }	16/16	100%	100%

Ahha, UG must be either the bistratal or monostratal grammars!

But as we saw from the state transition digram, the learner isn't actually learning very much.

Simulation 2

- s/z alternation of English plural suffix:
  - Two new constraints:
    - *voice: No voiced consonants.
    - *C̥C̬: No unvoiced consonant before a voiced consonant.
  - Voiced underlyingly
  - New base forms added to inventory

Lombardi (1996)

Simulation 2 Methods

- 8! = 40,320 fully stratified grammars
- 54 are ‘functionally’ distinct

“Extensional” Success

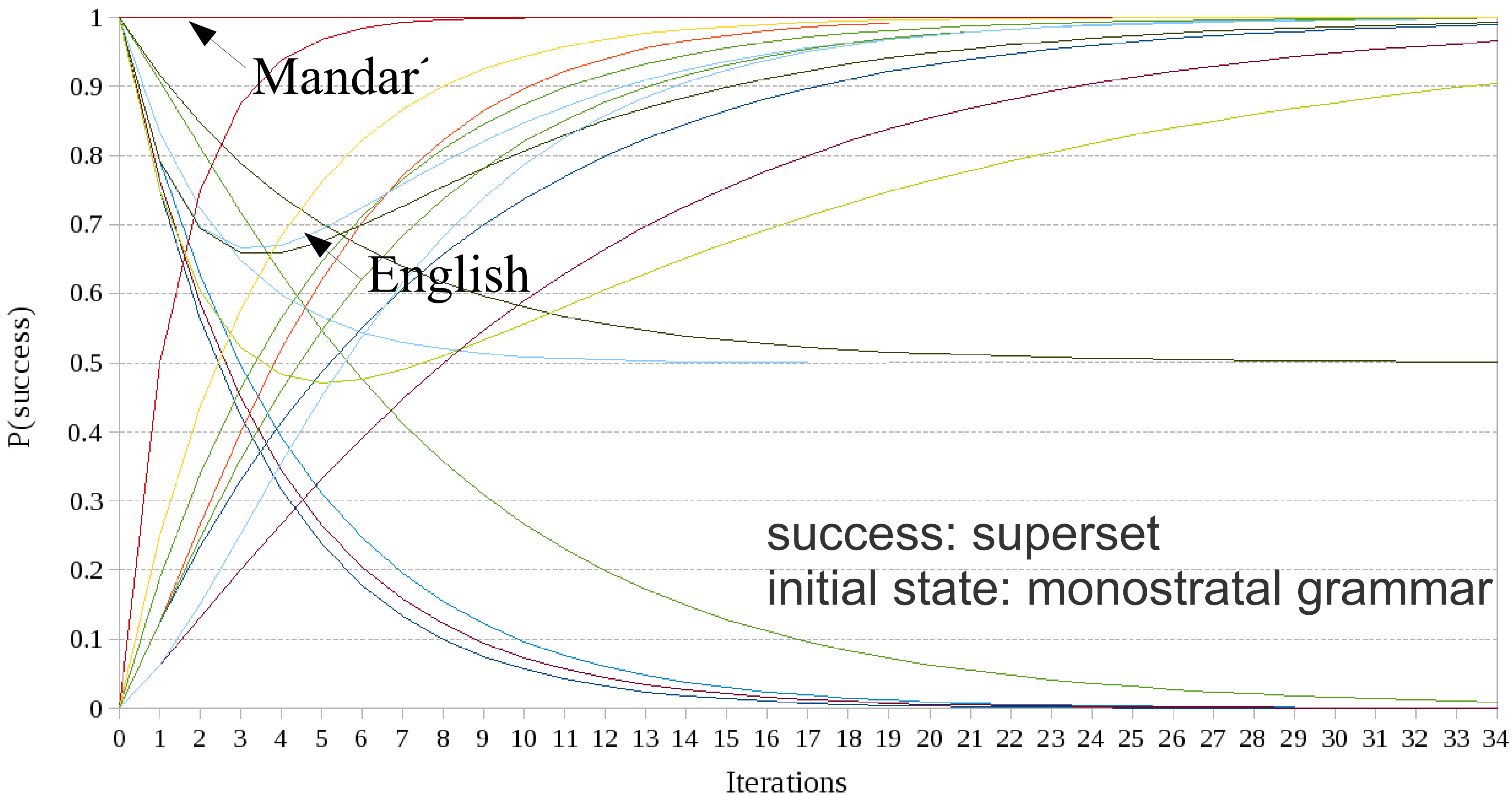
UG	Lx. Learnable	P(English Learned)	P(Mandar´ Learned)
M » F (720)	7/54	8%	0%
{M} » {F}	19/54	0%	0%
{ all }	2/54	0%	0%

“Superset” Success

UG	Lx. Learnable	P(English Learned)	P(Mandar´ Learned)
M » F (720)	24/54	86%	66%
{M} » {F}	44/54	100%	100%
{ all }	46/54	100%	100%

Even with the most generous notion of success, some languages still aren't learnable (but English and Mandar´ are OK).

Rate of Convergence



Conclusion

- How do we fairly choose a notion of “success” when evaluate learning algorithms?
  - “Extensional Success” too strong!
  - But “Superset Success” too weak!
- Are the unlearnable languages actual languages?
  - If not, this would indicate EDCD is actually on the right track.
- How do the alternative algorithms differ?

Denasalization

pæ̃nt	*NC̥	ID(voice)	ID$_C$(nas)
pæ̃nt	*!		
pæ̃nd		*!	
▶ pæ̃tt			*
pæ̃td		*!	*

English Plural Alternation

	pænt+z	*C̥C̬	ID(voice)	*voice
	pæntz	*!		*
▶	pænts		*	
	pændz		*	*!*
	pænds		**!	*

Can't Demote Faith. Constraints

Goal: *V$_{ORAL}$N ≫ *V$_{NASAL}$ ≫ ID(nasal)
Hypoth: ID(nasal) ≫ *V$_{NASAL}$ ≫ *V$_{ORAL}$N

Learner hears: /pænt/ → “pæ̃nt”

pæ̃nt	ID(nas)	*V$_{NASAL}$	*V$_{ORAL}$N
pænt	*!		*
► pæ̃nt		*	

Child's Form = Adult Form

Tesar & Smolensky (2000), Smith (2000)

Markov Model Details

- D^t an nxn matrix:
  - D^t_{ij} the probability that a learner who started in state i is in state j at time t.
  - $D^{t+1} = D^t M$
  - $D^0 = I$ (a learner who started in state i has, at time 0, probability 1.0 of being in state i)
  - $D^k = IM^k$

Superset Success