Computation, learning, and phonological typology

Class 1: Introduction

Adam Albright and Eric Baković

CreteLing 2023 — July 2023

creteling2023.phonology.party

Setting the stage

The goal of generative phonology

The goal of generative phonology

- Theory of possible vs. impossible human phonologies
  - A theory of typology
  - A theory of learnability

The goal of generative phonology

- Theory of possible vs. impossible human phonologies
  - A theory of typology
  - A theory of learnability
- Data: attested vs. unattested human phonologies
  - Typological description of language
  - Better: ‘verified typology’
  - Inferring what is (im)possible from what is (un)attested: a statistical problem

A possible statistical reformulation

- Frequent vs. rare vs. unattested phonologies
- Frequent = preferred in some way, rare and unattested = dispreferred/impossible

(Linear separation vs. modeling the observed distribution)

Assessing progress

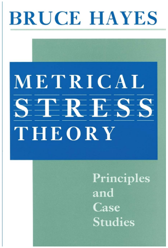
If our goal is to build a model that rules in what's possible, and rules out what's impossible, then we need…

- A way to calculate the predicted typology of the model
- A way to assess fit to empirical typology

Both have generally been approached rather informally, especially prior to Optimality Theory (OT; Prince & Smolensky 2004[1993])

Calculating predicted typologies: a pre-OT example

- Hayes (1995): proposed set of parameters for stress systems
- Predicted typology = step through +/– values for each parameter, selected combinations
- Focus: all known systems can be described with a combination of proposed parameter settings



- A somewhat unusual case! Not many 'complete parameter sets' for significant portions of phonological grammar

Calculating typologies in OT

Progress, but not solved

- For reasons that we'll discuss in detail, Optimality Theory has proven to be a useful framework for calculating typologies

Calculating typologies in OT

Gordon (2002) on quantity-insensitive stress

- Proposed set of 12 constraints for quantity-insensitive stress systems
  - System formally generates 152 possible patterns (or 302, depending on certain assumptions)
  - All 24 attested patterns are generated
  - . . . but 128 unattested patterns are also generated

Calculating typologies in OT

Kager (2012) on stress windows

- Compared various sets of $\approx$10 constraints for generating stress windows at word edges
  - Best system formally generates 98 distinct windows
    - All 22 attested patterns are generated
    - . . . but 76 unattested patterns are also generated (6/14 ‘default’ patterns, 70/84 lexical accent patterns)

A recurring theme

- All proposed theories on the market overgenerate
- Accidental gaps
  - Limited set of existing languages
  - Limited descriptions of existing languages
- Non-accidental, non-linguistic forces
  - Genetic relatedness, contact further limit diversity

A recurring theme (*cont.*)

- Linguistic factors

A recurring theme (*cont.*)

- Linguistic factors
    - Some patterns easier or harder to identify based on naturally occurring data?
    - Some patterns dispreferred because of properties of the grammars that generate them?
    - Some patterns dispreferred due to how humans learn grammars?
    - Some patterns dispreferred due to substantive biases?

Some goals of this class

- Introduce some concepts and tools that have proven useful in reasoning about the typologies that grammatical formalisms generate
- Incorporating learnability into models of typology
    - Typology shaped by the data available to learners
    - Learning biases (evaluation metrics) favor some patterns over others
    - Survey various proposed biases and consider their consequences for typology
- Examples drawn from phonology, largely using Optimality Theory, but concepts are widely applicable

What this class will not deliver

- Set of packaged results showing dramatically improved fits to attested typology
  - This is the kind of work we are hoping to inspire!
- Settled arguments in favor of particular learning models, biases, etc.

Using grammar to define a typology

Reminder: the challenge

- Take grammatical formalisms, and generate predicted typologies from them
- Assess fit to observed typology

We will focus mainly on the first task...

Grammar as parameter settings

- Models of grammar are sets of parameters
- Settings of parameters $\Rightarrow$ possible languages
  - Small (light switch) or infinite (lexicon)
  - Discrete or continuous

Parameters of rule-based approaches

The Sound Pattern of English (SPE; Chomsky & Halle 1968)

- Lexicon: strings of phonological segments, represented using phonological features
- Rules: A $\rightarrow$ B / C ___ D
  - Context and change = strings, represented using phonological features, boundary symbols, etc.
- Various properties of rules (iterative, direction of application, etc.)
- Rule ordering: sometimes $Q(P(x)) \neq P(Q(x))$

Generating a typology

Generative models

- Model of how the hypothesis (grammar) is constructed
- Generating underlying reprsentations (URs) for lexical items
- Generating rules
- Generating rule orders, etc.

Generating a very simple typology

A toy example

- Lexicon: just /sa/, /ʃa/, /si/, /ʃi/
  - Pick a number of lexical items
  - For each, pick [±anterior], pick [±high]
  - Remaining values are predictable, yielding 4 URs
- Rules: change anteriority of stridents, possibly conditioned by a vowel
  - Pick a number of rules
  - For each, pick a change direction: [αanterior]
  - Pick whether a context is specified
  - If yes, pick vowel feature [βhigh]
  - [+strid] $\rightarrow$ [αanterior] / ___ ([βhigh])

Some toy grammars

Lg 1
- Lexicon: /sa/, /ʃi/
- Rules: none

Lg 2
- Lexicon: /sa/, /ʃi/
- Rules: [+strid] $\rightarrow$ [–ant] / ___ [+high] (s $\longrightarrow$ ʃ / ___ i)

Lg 3
- Lexicon: /sa/, /ʃi/
- Rules: [+strid] $\rightarrow$ [+ant] / ___ [–high] (ʃ $\longrightarrow$ s / ___ a)

Lg 4
- Lexicon: /sa/, /ʃa/, /si/, /ʃi/
- Rules: [+strid] $\rightarrow$ [–ant] / ___ [+high] (s $\longrightarrow$ ʃ / ___ i)

…etc.

The contrast/neutralization typology

Observation: Given a potential opposition between two minimally distinct phones, ‘unmarked’ A (here, [s]) and ‘marked’ B ([ʃ]), there appear to be four basic possible distribution types.

1. **Full Contrast (FC):**
   [s] and [ʃ] contrast everywhere.
2. **Absolute Neutralization (AN):**
   Only [s] is ever found.
3. **Contextual Neutralization (CN):**
   Only [ʃ] is found in certain contexts (here, before [i]); elsewhere, [s] and [ʃ] contrast.
4. **Complementary Distribution (CD):**
   Only [ʃ] is found before [i]; elsewhere, only [s] is found.

Analysis in terms of (ordered) rules (= SPE)

'___ *i*' represents the palatalizing context potentially favoring [ʃ];

'___ *a*' represents the 'elsewhere' context potentially favoring [s].

Type	*Lexicon*	*Rule(s)*	*Derivations*				
1. **FC**	/si, ʃi, sa, ʃa/	—	—				
2a. **AN-a**	/si, sa/	—	—				
2b. **AN-b**	/si, ʃi, sa, ʃa/	ʃ ⟶ s	UR	/si/	/ʃi/	/sa/	/ʃa/
			ʃ ⟶ s	—	|si|	—	|sa|
			SR	[si]	[si]	[sa]	[sa]
3. **CN**	/si, ʃi, sa, ʃa/	s ⟶ ʃ / ___ i	UR	/si/	/ʃi/	/sa/	/ʃa/
			s ⟶ ʃ / ___ i	|ʃi|	—	—	—
			SR	[ʃi]	[ʃi]	[sa]	[ʃa]

Analysis in terms of (ordered) rules (= SPE)

'___ *i*' represents the palatalizing context potentially favoring [ʃ];

'___ *a*' represents the 'elsewhere' context potentially favoring [s].

Type	*Lexicon*	*Rule(s)*	*Derivations*				
4a. **CD-a**	/si, sa/	s ⟶ ʃ / ___ i	UR	/si/	—	/sa/	—
			s ⟶ ʃ / ___ i	|ʃi|	—	—	—
			SR	[ʃi]	—	[sa]	—
4b. **CD-b**	/si, ʃi, sa, ʃa/	ʃ ⟶ s s ⟶ ʃ / ___ i	UR	/si/	/ʃi/	/sa/	/ʃa/
			ʃ ⟶ s	|si|	|si|	|sa|	|sa|
			s ⟶ ʃ / ___ i	|ʃi|	|ʃi|	—	—
			SR	[ʃi]	[ʃi]	[sa]	[sa]

Parameters in Optimality Theory (OT)

- Lexicon: strings of phonological segments, represented using phonological features
- Constraints: convenient early assumption that these are fixed, universal (Prince & Smolensky, 2004[1993])
- Ranking of constraints: assumed to be total
  - Minimum: for k constraints, k–1 rankings A $\gg$ B

(Tesar & Smolensky, 2000)

Marginalizing over the lexicon

- Languages that we observe have both lexicons and grammars
- When evaluating a theory of rules/constraints, it is useful to know what languages it could generate *in principle*
  - i.e., independent of what the lexicon happens to be
- Approach: consider the set of outputs that the grammar would favor for *all possible inputs*
  - This is called *Richness of the Base* in OT

Constraint ranking as a parameter

Consider just two constraints

- $\mathbb{M}$arkedness: *[+strident]
  - Penalizes any occurrence of a [+strident] segment
  - Obeyed in Māori, Hawaiian, most Australian languages
- $\mathbb{F}$aithfulness: IDENT([±strident])
  - Penalizes any change of stridency between input and output; i.e., *[+strid]$\leftrightarrow$[–strid], or *ʃ$\leftrightarrow$h

Different rankings, different outcomes

- Two constraints ⇒ two rankings
- English: stridents may occur everywhere

/ʃ/	𝔽:IDENT([±strident])	𝕄:*[+strident]
☞ [ʃ]	✔	*
[h]	* W	✔ L

- Māori: no stridents anywhere

/ʃ/	𝕄:*[+strident]	𝔽:IDENT([±strident])
[ʃ]	* W	✔ L
☞ [h]	✔	*

Enlarging the typology

Adding one additional constraint

- $\mathbb{M}^{g}$: *[+strident]
  - Penalizes any occurrence of a [+strident] segment
- $\mathbb{M}^{s}$: *[+strident]#
  - Penalizes any *word-final* [+strident] segment
  - Obeyed in Caribbean Spanish, Thai, Korean, Vietnamese
  - More accurately: no stridents *unless before V* (*sC, *s#)
- $\mathbb{F}$: IDENT([±strident]): penalizes any change of stridency between input and output

3 constraints $\Rightarrow$ 3! rankings = 6 (FACTORIAL TYPOLOGY)

Faithfulness rules: $\mathbb{F} \gg \{\mathbb{M}^g, \mathbb{M}^s\}$

/sæp/	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^g$:*[+str]	$\mathbb{M}^s$:*[+str]#
☞ [sæp]	✔	*	✔
[hæp]	* W	✔ L	✔

/pæs/	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^g$:*[+str]	$\mathbb{M}^s$:*[+str]#
☞ [pæs]	✔	*	*
[pæh]	* W	✔ L	✔ L

– $\mathbb{F}$:IDENT([±strid]) at the top protects all stridents

– Stridents & non-stridents are contrastive in all positions

(NB. 2 of the 6 rankings)

Faithfulness tempered: $\mathbb{M}^s \gg \mathbb{F} \gg \mathbb{M}^g$

/sæp/	$\mathbb{M}^s$:*[+str]#	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^g$:*[+str]
☞ [sæp]	✔	✔	*
[hæp]	✔	* W	✔ L

/pæs/	$\mathbb{M}^s$:*[+str]#	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^g$:*[+str]
[pæs]	* W	✔ L	*
☞ [pæh]	✔	*	✔

– Contextual $\mathbb{M}^s$:*[+str]# bans final stridents, but $\mathbb{F}$:IDENT([±str]) protects stridents elsewhere

– Stridents are contrastive only in non-final position

(NB. 1 of the 6 rankings)

Faithfulness lost: $\mathbb{M}^g \gg \mathbb{F}$ ($\mathbb{M}^s$ is *unrankable*)

/sæp/	$\mathbb{M}^g$:*[+str]	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^s$:*[+str]#
☞ [hæp]		*	
[sæp]	*W	L	

/pæh/	$\mathbb{M}^g$:*[+str]	$\mathbb{F}$:IDENT([±str])	$\mathbb{M}^s$:*[+str]#
☞ [hæp]		*	
[pæs]	*W	L	*W

- $\mathbb{M}^g$:*[+strid] bans all stridents
- Stridents are never contrastive with non-stridents

(NB. remaining 3 of 6 rankings)

Factorial typologies

- k constraints yields at most $k!$ different languages
- In practice, usually many fewer, once identical sets of outputs are grouped together
- Some patterns may be generated by many rankings, some by few or even just one
- See McCarthy (2008), Ch. 5 for further examples and discussion (uploaded as part of Background Readings)

Typology depends on CON

Naturally, changing CON affects the resulting typology

Recall the contrast/neutralization typology

1. **Full Contrast (FC):**
   [s] and [ʃ] contrast everywhere.
2. **Absolute Neutralization (AN):**
   Only [s] is ever found.
3. **Contextual Neutralization (CN):**
   Only [ʃ] is found in certain contexts (here, before [i]); elsewhere, [s] and [ʃ] contrast.
4. **Complementary Distribution (CD):**
   Only [ʃ] is found before [i]; elsewhere, only [s] is found.

Analysis in terms of ranked constraints (= OT)

- **Constraints:** ʃ $\longrightarrow$ s = $\mathbb{M}^{g}$:*ʃ $\gg$ $\mathbb{F}$:*s↔ʃ

  s $\longrightarrow$ ʃ / __ i = si $\longrightarrow$ ʃi = $\mathbb{M}^{s}$:**si* $\gg$ $\mathbb{F}$:*s↔ʃ
- **3 constraints:** $3! = 3 \times 2 \times 1 = 6$ total rankings
- **Inputs:** /si/, /ʃi/, /sa/, /ʃa/
- **Candidate outputs:** [si], [ʃi], [sa], [ʃa] (NB. only *s*↔*ʃ* changes)
- **Unfaithful input-output mappings** are marked with underlining, e.g. /si/ $\longrightarrow$ [ʃi]

Full contrast (FC): $\mathbb{F} \gg \{\mathbb{M}^s, \mathbb{M}^g\}$

/si/	$\mathbb{F}$:*$s \leftrightarrow ʃ$	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*$ʃ$
☞ [si]		*	
[ʃ̠i]	*W	L	*W

/ʃi/	$\mathbb{F}$:*$s \leftrightarrow ʃ$	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*$ʃ$
☞ [ʃi]			*
[s̠i]	*W	*W	L

/sa/	$\mathbb{F}$:*$s \leftrightarrow ʃ$	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*$ʃ$
☞ [sa]			
[ʃ̠a]	*W		*W

/ʃa/	$\mathbb{F}$:*$s \leftrightarrow ʃ$	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*$ʃ$
☞ [ʃa]			*
[s̠a]	*W		L

(NB. 2 of 6 total rankings)

Absolute neutralization (AN): $\mathbb{M}^g \gg \{\mathbb{M}^s, \mathbb{F}\}$

/si/	$\mathbb{M}^g$:*ʃ	$\mathbb{M}^s$:*si	$\mathbb{F}$:*s↔ʃ
☞ [si]		*	
[ʃ̲i]	*W	L	*W

/ʃi/	$\mathbb{M}^g$:*ʃ	$\mathbb{M}^s$:*si	$\mathbb{F}$:*s↔ʃ
☞ [s̲i]		*	*
[ʃi]	*W	L	L

/sa/	$\mathbb{M}^g$:*ʃ	$\mathbb{M}^s$:*si	$\mathbb{F}$:*s↔ʃ
☞ [sa]			
[ʃ̲a]	*W		*W

/ʃa/	$\mathbb{M}^s$:*ʃ	$\mathbb{M}^g$:*si	$\mathbb{F}$:*s↔ʃ
☞ [s̲a]			*
[ʃa]	*W		L

(NB. 2 of 6 total rankings)

Contextual neutralization (CN): $\mathbb{M}^s \gg \mathbb{F} \gg \mathbb{M}^g$

/si/	$\mathbb{M}^s$:**si*	$\mathbb{F}$:**s*↔*ʃ*	$\mathbb{M}^g$:**ʃ*
☞ [ʃ̱i]		*	*
[si]	*W	L	L

/ʃi/	$\mathbb{M}^s$:**si*	$\mathbb{F}$:**s*↔*ʃ*	$\mathbb{M}^g$:**ʃ*
☞ [ʃi]			*
[s̱i]	*W	*W	L

/sa/	$\mathbb{M}^s$:**si*	$\mathbb{F}$:**s*↔*ʃ*	$\mathbb{M}^g$:**ʃ*
☞ [sa]			
[ʃ̱a]		*W	*W

/ʃa/	$\mathbb{M}^s$:**si*	$\mathbb{F}$:**s*↔*ʃ*	$\mathbb{M}^g$:**ʃ*
☞ [ʃa]			*
[s̱a]		*W	L

(NB. 1 of 6 total rankings)

Complementary distribution (CD): $\mathbb{M}^s \gg \mathbb{M}^g \gg \mathbb{F}$

/si/	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*ʃ	$\mathbb{F}$:*s↔ʃ
☞ [ʃ̠i]		*	*
[si]	*W	L	L

/ʃi/	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*ʃ	$\mathbb{F}$:*s↔ʃ
☞ [ʃi]		*	
[s̠i]	*W	L	*W

/sa/	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*ʃ	$\mathbb{F}$:*s↔ʃ
☞ [sa]			
[ʃ̠a]		*W	*W

/ʃa/	$\mathbb{M}^s$:*si	$\mathbb{M}^g$:*ʃ	$\mathbb{F}$:*s↔ʃ
☞ [s̠a]			*
[ʃa]		*W	L

(NB. remaining 1 of 6 total rankings)

An alternative constraint set

- **Constraints:** $\int \longrightarrow$ s $=$ $\mathbb{M}^{g}$:*$ʃ$ $\gg$ $\mathbb{F}^{u}$:*$ʃ\!\to\! s$
  s $\longrightarrow$ ʃ / __ i $=$ si $\longrightarrow$ ʃi $=$ $\mathbb{M}^{s}$:*si $\gg$ $\mathbb{F}^{m}$:*$s\!\to\!ʃ$
- **4 constraints:** $4! = 4 \times 3 \times 2 \times 1 = 24$ total rankings
- **Inputs:** /si/, /ʃi/, /sa/, /ʃa/
- **Candidate outputs:** [si], [ʃi], [sa], [ʃa] (NB. only *s*↔*ʃ* changes)
- **Unfaithful input-output mappings** are marked with underlining, e.g. /si/ $\longrightarrow$ [ʃi]

FC: $\mathbb{F}^u \gg \mathbb{M}^g \gg \mathbb{M}^s$ ***or*** $\{(\mathbb{F}^m \gg \mathbb{M}^s)$ **&** $(\mathbb{F}^u \gg \mathbb{M}^g)\}$

/si/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^g$ *ʃ	$\mathbb{M}^s$ *si
☞ [si]				*
[ʃ̲i]	*W		*W	L

/ʃi/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^g$ *ʃ	$\mathbb{M}^s$ *si
☞ [ʃi]			*	
[s̲i]		*W	L	*W

/sa/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^g$ *ʃ	$\mathbb{M}^s$ *si
☞ [sa]				
[ʃ̲a]	*W		*W	

/ʃa/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^g$ *ʃ	$\mathbb{M}^s$ *si
☞ [ʃa]			*	
[s̲a]		*W	L	

(NB. 8 of 24 total rankings)

AN: $\mathbb{M}^g \gg \{\mathbb{F}^u, \mathbb{M}^s\}$ ($\mathbb{F}^m$ is unrankable)

/si/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si
☞ [si]				*
[ʃi]	*W	*W		L

/ʃi/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si
☞ [si]			*	*
[ʃi]		*W	L	L

/sa/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si
☞ [sa]				
[ʃa]	*W	*W		

/ʃa/	$\mathbb{F}^m$ *s→ʃ	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si
☞ [sa]			*	
[ʃa]		*W	L	

(NB. 8 of 24 total rankings)

CN: $(\{\mathbb{F}^u, \mathbb{M}^s\} \gg \mathbb{M}^g)$ & $(\mathbb{M}^s \gg \mathbb{F}^m)$

/si/	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^m$ *s→ʃ
☞ [ʃi]			*	*
[s̲i]		*W	L	L

/ʃi/	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^m$ *s→ʃ
☞ [ʃi]			*	
[s̲i]	*W	*W	L	

/sa/	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^m$ *s→ʃ
☞ [sa]				
[ʃ̲a]			*W	*W

/ʃa/	$\mathbb{F}^u$ *ʃ→s	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^m$ *s→ʃ
☞ [ʃa]			*	
[s̲a]	*W		L	

(NB. 4 of 24 total rankings)

CD: $\left(\mathbb{M}^s \gg \mathbb{M}^g \gg \mathbb{F}^u\right)$ & $\left(\mathbb{M}^s \gg \mathbb{F}^m\right)$

/si/	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{F}^m$ *s→ʃ
☞ [ʃi]		*		*
[s̲i]	*W	L		L

/ʃi/	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{F}^m$ *s→ʃ
☞ [ʃi]		*		
[s̲i]	*W	L	*W	

/sa/	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{F}^m$ *s→ʃ
☞ [sa]				
[ʃ̲a]		*W		*W

/ʃa/	$\mathbb{M}^s$ *si	$\mathbb{M}^g$ *ʃ	$\mathbb{F}^u$ *ʃ→s	$\mathbb{F}^m$ *s→ʃ
☞ [sa]			*	
[ʃ̲a]		*W	L	

(NB. 3 of 24 total rankings)

Reverse neutralization! (RN): $\mathbb{F}^m \gg \mathbb{M}^s \gg \mathbb{M}^g \gg \mathbb{F}^u$

/si/	$\mathbb{F}^m$ $*s{\to}ʃ$	$\mathbb{M}^s$ $*si$	$\mathbb{M}^g$ $*ʃ$	$\mathbb{F}^u$ $*ʃ{\to}s$
☞ [si]		*		
[ʃ̠i]	*W	L	*W	

/ʃi/	$\mathbb{F}^m$ $*s{\to}ʃ$	$\mathbb{M}^s$ $*si$	$\mathbb{M}^g$ $*ʃ$	$\mathbb{F}^u$ $*ʃ{\to}s$
☞ [ʃi]			*	
[s̠i]		*W	L	*W

/sa/	$\mathbb{F}^m$ $*s{\to}ʃ$	$\mathbb{M}^s$ $*si$	$\mathbb{M}^g$ $*ʃ$	$\mathbb{F}^u$ $*ʃ{\to}s$
☞ [sa]				
[ʃ̠a]	*W		*W	

/ʃa/	$\mathbb{F}^m$ $*s{\to}ʃ$	$\mathbb{M}^s$ $*si$	$\mathbb{M}^g$ $*ʃ$	$\mathbb{F}^u$ $*ʃ{\to}s$
☞ [sa]				*
[ʃ̠a]			*W	L

(NB. remaining 1 of 24 total rankings)

Factorial typology and universal CON

- Ability to enumerate entire predicted typology was an interesting selling point for OT
- Useful assumption in guiding formulation of new constraints
- Fixed universal CON made calculating typology easy and accessible*

*But it's still *very* challenging to do by hand! Software helps.

Relaxing these assumptions

- Finite set, but constraints can be active/inactive
  - Expands number of possible rankings considerably
  - Sum of factorials, but still finite
- Infinite constraint set
  - Can’t be enumerated, but can be sampled (similar to SPE rules)
  - Sampling requires a probability distribution over rules/constraints, such as that assigned by a generative model
- Next: how properties of rules/constraints may affect their probability

References

CHOMSKY, NOAM, and MORRIS HALLE. 1968. *The Sound Pattern of English*. New York: Harper and Row.

GORDON, MATTHEW. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20.491–552.

HAYES, BRUCE. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: The University of Chicago Press.

KAGER, RENÉ. 2012. Stress in windows: Language typology and factorial typology. *Lingua* 122.1454–1493.

MCCARTHY, JOHN J. 2008. *Doing Optimality Theory*. Malden, MA: Wiley-Blackwell.

PRINCE, ALAN, and PAUL SMOLENSKY. 2004[1993]. *Optimality theory: Constraint interaction in generative grammar*. Malden: Blackwell. Available on ROA (http://roa.rutgers.edu).

TESAR, BRUCE, and PAUL SMOLENSKY. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.