

# Phonology looking forward

Michael Hammond

U. of Arizona

Aug. 5, 2024

# The plan today

What's the plan today?

- ▶ Basic structure of phonology
- ▶ Dealing with opacity in phonology
- ▶ Combining neural and symbolic approaches to phonology

# Overview

Overview

History

Rules and constraints

Opacity

Strata, cophonologies, and grammars

The future

Neuro-symbolic computation

Conclusion

References

# What is phonology?

The majority of human languages are spoken languages where words and sentences are made up of sounds. These sounds are subject to at least these important constraints:

- ▶ The set of sounds for any language is finite.
- ▶ The sounds themselves are meaningless.
- ▶ The set of sounds a language has is drawn from a larger (finite) set of possible sounds.
- ▶ The set of sounds any language has is not random.
- ▶ The cooccurrence of sounds in words and phrases in any language is also not random.

Why?

# Finiteness

Very rough counts of contrastive sounds in various languages.

Hawaiian	13
Persian	29
Ossetian	32-33
Sorani	35-37
Pashto	41
English	44
!Xõ	160

The upshot: there's a **lot** of variation.

# Are segments randomly selected?

This is a wonderful website where you can look at phonemic segment inventories for thousands of languages:

► <https://phoible.org/>

Two kinds of obvious generalizations:

1. There are contingency relationships between segments (Greenbergian universals).
2. Languages exploit the same dimensions: “birds of a feather flock together”.

# Contingency relationships

The presence of a segment depends on the presence of another segment. For example:

- ▶ If a language has [d], it will also have [t].
- ▶ If a language has a nasal vowel like [ã], then it will have the corresponding oral vowel [a].
- ▶ If a language has ejective consonants, then it will have the non-ejective counterparts.
- ▶ Etc.

(Greenberg, 1966, 1978, ...)

## Dimensions of inventories: Pashto

In an inventory, segments tend to appear in the same rows and columns, where those rows and columns are articulatory dimensions.

	lab	den	pal-alv	retro	pal	vel	uvul	glot
nas	m	n		ɳ		ŋ		
stop	p,b	t,d		ʈ,ɖ		k,g	(q)	
affric		ʈʂ,ɖʂ	ʈʃ,ɖʃ					
fric	(f)	s,z	ʃ,ʒ	ʂ,ʐ		x,ɣ		
approx		l,r		ɽ	j	w		h

For example, Pashto has a whole series of retroflex consonants, **not just one**.



# Testing “birds of a feather”

- ▶ We download inventory data from the phoible website, specifically the UPSID subset.
- ▶ We count up how often each segment occurs.
- ▶ We look at whether segments are independent or whether they are more likely to cooccur.

(prolog program does all this: `segs.pl`)

# Cooccurrence testing

seg	frequency	seg	frequency	actual joint freq	expected
t	34/451 = 0.075	d	27/451 = 0.059	21/451 = 0.046	0.004
x	94/451 = 0.208	γ	55/451 = 0.121	28/451 = 0.062	0.025
η	24/451 = 0.053	η	237/451 = 0.525	21/451 = 0.046	0.027
t	181/451 = 0.401	d	27/451 = 0.059	3/451 = 0.006	0.024
η	24/451 = 0.053	k	403/451 = 0.893	22/451 = 0.048	0.047
d	120/451 = 0.266	x	94/451 = 0.208	22/451 = 0.048	0.055

# Phonotactics

The sounds of a language are not randomly distributed in words or phrases. For example:

- ▶ English and Persian both have [s] and [t], but English allows [st] at the beginning of words, but Persian does not.
- ▶ English and Persian both have [b] and [r], but Persian allows [br] at the end of a word, e.g. in [abr] ‘cloud’, and English does not.

There's lots of variation in this domain as well.

# Calculating phonotactics

We can calculate the average predictability of segments based on previous segments in a language using **average entropy**.

$$H(X) = -\frac{1}{n} \sum_{i=2}^n p(x_i|x_{i-1}) \log_2 p(x_i|x_{i-1})$$

# Average entropy for some languages

How predictable is a segment given the preceding segment?

1 = not predictable at all

0 = completely predictable

Tajik	0.27
Pashto	0.25
Northern Kurdish	0.27
Central Kurdish	0.26
Persian	0.26

These languages have very different inventories and phonotactics, but these numbers are strikingly similar.

(Data from <https://github.com/CUNY-CL/wikipron/>,  
prolog code in `ent.pl`)

# Phonological framework

Inventories and phonotactics are not random. How do we treat them? Historically, there have been two broad approaches (Anderson, 1985):

- ▶ rules
- ▶ representations/constraints

# Rules

Rule-based accounts have been fairly typical for morphophonemic alternations. Consider the imperative/subjunctive prefix *be-* in Persian. This assimilates to a following vowel in the following way.

Ci	be-gir, bi-gir	‘get!’
Cu	be-gu, bu-gu, bo-gu	‘say!’
Ce	be-keʃ	‘pull!’
Co	bo-xor	‘eat!’
Cæ	be-xær	‘buy!’
Ca	be-xab	‘sleep!’
V	bi-(j)a	‘come!’

# Persian vowel assimilation

Here is the pattern summarized:

- ▶ obligatory assimilation to [o] when the first stem vowel is [o]
- ▶ optional assimilation to [o] when the first stem vowel is [u]
- ▶ optional raising when the first stem vowel is high.
- ▶ the raising does **not** apply if rounding is possible, but doesn't happen; we do **not** get \*[bi-gu].
- ▶ if the stem begins with a vowel, we get [i] and an inserted glide.



## A partial rule-based analysis of the first generalization

The classical generative analysis is rule-based (Chomsky and Halle, 1968). The generalization is captured by transforming one vowel into another.

$$\begin{bmatrix} +\text{syl} \\ -\text{high} \\ -\text{low} \\ -\text{back} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{round} \\ +\text{back} \end{bmatrix} / \text{---} [-\text{syl}] \begin{bmatrix} +\text{round} \\ -\text{high} \end{bmatrix}$$

(The vowel [e] becomes [o] when it's followed by a consonant and [o].)

## How rule-based phonology worked

- ▶ The theory claimed that the analysis of any language was a finite **ordered** sequence of rules, e.g.  $R_1, R_2, \dots, R_n$ .
- ▶ Rules had to be written using a specific formal system. That requirement embodied a claim about what could be a rule.
- ▶ Rules had to use specific phonological features, specific phonetic properties that rules could refer to **and no others**. Again, this reflected a claim that only certain phonetic dimensions were available in rules.
- ▶ The theory included a specific algorithm for choosing between analyses: **the evaluation metric**.

## Evaluation metric

- ▶ Choose an analysis with fewer rules.
- ▶ Choose an analysis with less complex rules.

There was no explicit algorithm for figuring out the rules from the data, only a method for choosing between alternate analyses.

## Rule application

Rules are **independent** and apply in sequence. Their effects can be undone or made opaque by later rules in the sequence. For example, Persian has two other rules that interact with the rule above:

- ▶ Insert an [e] between an obstruent and a liquid word-initially.
- ▶ Delete a vowel in a double-sided open syllable (optional).

This gives us interactions like this:

input	profesor
vowel insertion	perofesor
vowel harmony	porofesor
vowel deletion	porfesor

Notice how the vowel that triggers the harmony is **not** evident in the surface form *porfesor*.

# What's wrong with rules?

- ▶ There are an infinite number of possible rules.
- ▶ The number of rules an analysis might contain is unbounded.
- ▶ There is no accepted algorithm for finding the rules from data.
- ▶ At least some rules seem to apply in a **conspiratorial** way.

# Cluster conspiracy in Yawelmani

Consider these two rules in Yawelmani:

Insert a vowel to break up a three-consonant cluster:

$$\emptyset \rightarrow V / C\_CC$$

Delete a morpheme-initial consonant after two other consonants:

$$C \rightarrow \emptyset / CC + \_$$

These are separate rules in the phonology, but they achieve similar effects: eliminating three-consonant clusters (Kisseberth, 1970).

# Stress conspiracy in English

There are also conspiracies involving stress in English:

- ▶ Stress falls on every other syllable counting from the right, e.g. in *Àpalàchicóla* or *hàmamèlidáanthemum*.
- ▶ Stress shifts in adjectives to avoid adjacent stresses, e.g. in *thìrtéen* vs. *thírtèen mén*.
- ▶ Stressless vowels are deleted unless that would bring two stresses together, e.g. *opera* [ápəɾə] → [áprə], but *operatic* [ápəɾæɾík], \*[àpræɾík].

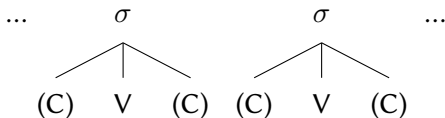
# Conspiracy as structure

- ▶ Build an **invisible** structure at some point in the derivation.
- ▶ That structure perseveres through all or some subsequent stages.
- ▶ That structure enforces whatever **global** constraints you need.



## Some conspiracies follow from syllable structure

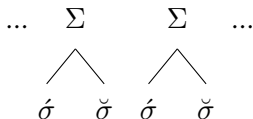
Let's imagine that a word is composed of a sequence of syllables and that a syllable is composed of a vowel with an optional consonant on each side.



If this characterization holds of the entire phonological derivation, it follows that we cannot have three-consonant clusters, because there is no way to parse them into syllables.

## Some conspiracies follow from metrical structure

Let's imagine that a foot is a stressed syllable followed by a stressless syllable and that English words must be parsed into feet.



If this characterization holds of the entire phonological derivation, it follows that we cannot have adjacent stresses.

# Are there non-prosodic conspiracies?

Do we get non-prosodic conspiracies? What would that look like?

We've seen that Persian [e] becomes [o] when the following vowel is [o]. Is [eCo] generally avoided? Is that restriction satisfied in other ways in other contexts?

This kind of thing doesn't seem to happen.

**Hypothesis:** all conspiracies involve structures that persevere through some or all of a derivation.

# Optimality Theory

Optimality Theory (Prince and Smolensky, 1993; McCarthy and Prince, 1993) posits constraints that apply across an entire derivation. This accounts for conspiracy effects directly.

However, this makes the incorrect prediction that we should see conspiratorial behavior even with non-prosodic phonology.

# Optimality Theory

- ▶ There is a finite set of universal constraints.
- ▶ Languages differ only in the **ranking** of constraints.
- ▶ A derivation involves matching the input with **every possible output form** and letting the constraints choose the best output.

# Persian harmony


Here's the general logic of an OT analysis of Persian vowel assimilation:

- ▶ We have a constraint against the output sequence [...eCo...].
- ▶ We have any number of ways we might avoid that sequence and we posit constraints **against** all of them.
- ▶ The constraint against converting the [e] to [o] is ranked **lowest**, making this the least painful way to avoid [...eCo...].

# A schematic tableau

Ranking is indicated with left-to-right order.

Constraints ruling out other repair strategies are indicated with “...” here.



/eCo/	*eCo	...	*e → o
oCo			*
eCo	*!		
eCe		*!	
eCCo		*!	
...		*!	

# The opacity problem

What about **porfesor**?

- ▶ The first vowel harmonizes with a vowel that does not surface!
- ▶ The alternative pronunciation \*perfeser should be legal as it does not violate the \*eCo constraint.



# Approaches to opacity

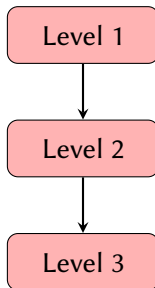
This is a very long-standing problem in Optimality Theory and a number of theories address it in various ways.

- ▶ Output-output correspondence (Kenstowicz, 1995)
- ▶ Sympathy theory (McCarthy, 1998)
- ▶ Stratal OT (Kiparsky, 2000, 2015)
- ▶ Cophonologies (Inkelas et al., 1996; Inkelas and Zoll, 2007)

Let's develop an answer to the *porfesor* problem that builds on both Stratal OT and Cophonologies.

# Stratal OT

The basic idea behind Stratal OT is that there are mini-OT phonologies associated with blocks of morphological processes.



There should then be a clear morphological demarcation between each block.

This doesn't help us with Persian since no morphology is involved.

# Cophonologies

*A cophonology ... is a phonological grammar, i.e. an input-output mapping, which coexists with other phonological grammars in a single language (Inkelas et al., 1996).*

This is used to capture generalizations that happen to only part of a language.

We'll use this to capture phonological relationships that cannot be modeled with orthodox OT.

## An alternative

- ▶ A derivation is an arrangement of **grammars**.
- ▶ Grammars are associated with arbitrary classes of morphemes.
- ▶ A word or phrase can be associated with multiple grammars.

Let's refer to a phonology organized like this as a **multi-grammar** or **multi-phonology**.

## A little more formally....

We can use this system in a number of ways:

- ▶ for exceptional behavior, like cophonologies:

$\text{stem}_{g_1}$  vs.  $\text{stem}_{g_2}$

- ▶ for morphological conditioning, like level ordering:

$\text{stem}_{g_1} + \text{sfx}_{g_2} + \text{sfx}_{g_3}$

- ▶ for opacity:

$\text{stem}_{g_1, g_2}$

Let's go through examples of each.

# Exceptional stress in English and Tajiki

stem<sub>g<sub>1</sub></sub> vs. stem<sub>g<sub>2</sub></sub>

## ► English

**Nouns** Rightmost stress falls on one of the last three syllables.

**Verbs** Rightmost stress falls on one of the last two syllables.

## ► Tajiki

**Nouns** Stress falls on the rightmost syllable.

**Verbs** Stress falls on the first syllable.

(See the talk later in the series by Mohsen Mahdavi on stress in various Iranian languages.)

# Morphological conditioning in Welsh

$$\text{stem}_{g_1} + \text{sfx}_{g_2} + \text{sfx}_{g_3}$$

The Welsh comparative and superlative suffixes trigger devoicing of a stem-final stop, but other suffixes do not.

stem	comparative	other
gwlyb 'wet'	gwly <b>p</b> ach 'wetter'	gwlybaniaeth 'moisture'
caled 'hard'	ca <b>l</b> etach 'harder'	caledu 'harden'
parod 'ready'	par <b>o</b> tach 'readier'	parodi 'to ready'
gwag 'empty'	gwa <b>c</b> ach 'more empty'	gwagedd 'vanity'
tebyg 'similar'	teby <b>c</b> ach 'more similar'	tebygu 'to assume'
cymreig 'Welsh'	cymrei <b>c</b> ach 'more Welsh'	cymreigaidd 'Welshlike'

# Opacity in Persian

stem <sub>$g_1, g_2$</sub>

- ▶  $g_1$ : loanwords  
profesor → porofesor
- ▶  $g_2$ : casual style  
porofesor → porfesor



## How does this differ?

- ▶ This is different from cophologies because a single item can appear in multiple sequential grammars.
- ▶ This is different from stratal ordering in that grammars are not necessarily ordered and are not necessarily associated with specific morphology.

(Lots of other issues to sort out here!)

# Traditional phonology and the future

- ▶ So far, we've been talking about traditional symbolic phonology, but perhaps this is out of date. Why?
- ▶ Rule-based formalisms are built on models of human cognition that were in vogue in the middle of the last century.
- ▶ In fact, constraint-based Optimality Theory was designed as an explicit compromise between symbolic approaches and neural nets.
- ▶ For concrete computational applications, modern neural nets perform much better than symbolic applications.

# Neural nets

Assume we want to have a neural net model the relationship between input and output forms.

1. Get a set of input–output pairs.
2. Convert those to numbers, e.g. convert each sound to a different number and represent each form as a vector of numbers.
3. Build a huge nested/sequential regression model that converts from one sequence of numbers to another.
4. Run the input–output pairs through the network again and again adjusting parameters for each batch of data until correct outputs are generated for each input.

# Why and why not neural nets

- ▶ Why?
  - ▶ It's provable that a sufficiently large neural net is "Turing-complete", that it can model any relationship we might want.
  - ▶ With sufficient appropriate data, a suitable network will **learn** whatever generalizations exist in the data.
- ▶ Why not?
  - ▶ Do we have sufficient training data?
  - ▶ The network can't tell us **what** the generalizations are.
  - ▶ A neural net can learn **anything**, including things that aren't languages.
- ▶ So do we want accuracy/precision or do we want interpretability?

Why not both?

# Neuro-symbolic computation

- ▶ The general idea behind neuro-symbolic computation is to **combine** symbolic and neural approaches (Valiant, 2008; Marcus, 2020).
- ▶ There have been various related ideas in linguistics, e.g. Prince and Smolensky (1993), Townsend and Bever (2001), Smolensky and Goldrick (2016).
- ▶ Optimality Theory was, in fact, a very restricted integration of neural/connectionist approaches and symbolic approaches. We can take OT with gradient constraints as moving OT even more toward a neural system, e.g. Legendre et al. (1990), Smolensky (2006), Hayes and Wilson (2008), etc.
- ▶ I'll take a different approach here (Winters et al., 2021; Huang et al., 2021).

# Combining symbolic and neural approaches

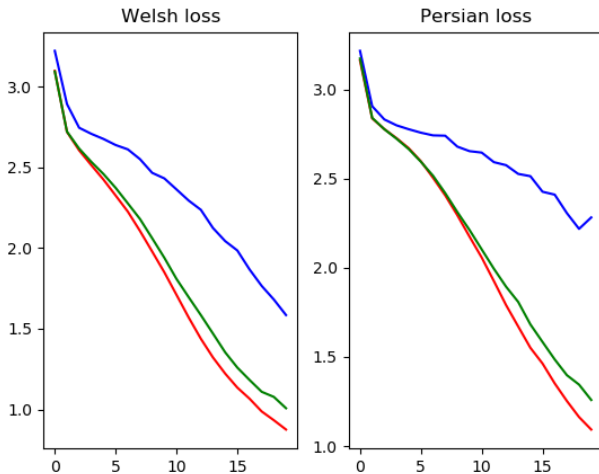
- ▶ One can add rules/constraints **on top of** a neural net.
- ▶ If the rules are correct, neural nets trained like this may perform better.
- ▶ Incorrect rules have the opposite effect.
- ▶ The associated rules are directly interpretable (though the rest of the net may not be).

# Neural nets and phonology

- ▶ Use data from Persian and Welsh.
- ▶ Build a neural net that predicts the next segment from previous segments.
- ▶ Add a rule on top of the network that rules out geminate consonants, an incorrect rule for both languages.
- ▶ Or, add a rule on top of the network that rules out word-initial geminates, a correct rule for both languages.

(python program does this: `neuro.py`)

# Results



(blue with gemination rule; green with initial gemination rule; red with no gemination rule)



# Neuro-symbolic results

- ▶ Adding a rule against geminates impairs learning for both Welsh and Persian, but more so for Persian.
- ▶ In Welsh, gemination occurs in  $\frac{54}{10250} = 0.005$  examples, but in Persian, gemination occurs in  $\frac{322}{9448} = 0.034$  examples.
- ▶ This numerical asymmetry correlates with the learning rate difference.

There are lots of ways to explore this further!

# Conclusions

What have we done?

- ▶ We've reviewed the general goals of phonology.
- ▶ We've talked about rule-based and constraint-based approaches.
- ▶ We've outlined a new approach to opacity in a constraint-based system: **multi-grammars**.
- ▶ We've outlined a simple approach to integrating symbolic analyses with neural approaches in modern phonology: **neurosymbolic phonology**.

# Selected references I

- Ambridge, B. and Blything, L. (2024). Large language models are better than theoretical linguists at theoretical linguistics. *Theoretical Linguistics*, 50:33–48.
- Anderson, S. R. (1985). *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. University of Chicago Press, Chicago.
- Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper & Row, New York.
- Greenberg, J. (1966). *Language universals, with specific reference to feature hierarchies*. Mouton, The Hague.
- Greenberg, J. H., editor (1978). *Universals of Human Language: Phonology*, volume 2. Stanford University Press, Stanford.
- Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39:379–440.
- Huang, J., Li, Z., Chen, B., Samel, K., Naik, M., Song, L., and Si, X. (2021). Scallop: From probabilistic deductive databases to scalable differentiable reasoning. *Advances in Neural Information Processing Systems*, 34:25134–25145.
- Inkelas, S., Orgun, C. O., and Zoll, C. (1996). Exceptions and static phonological patterns: cophonologies vs. prespecification. ROA-124, Rutgers Optimality Archive.
- Inkelas, S. and Zoll, C. (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics*, 45:133–171.
- Kenstowicz, M. (1995). Base-Identity and uniform exponence: Alternatives to cyclicity. Rutgers Optimality Archive.
- Kiparsky, P. (2000). Opacity and cyclicity. *Linguistic Review*, 17:1–15.
- Kiparsky, P. (2015). Stratal OT: A synopsis and FAQs. In Hsiao, Y. E. and Wee, L.-H., editors, *Capturing Phonological Shades within and across Languages*, pages 2–44. Cambridge Scholars Publishing, Newcastle upon Tyne.
- Kisseberth, C. (1970). On the functional unity of phonological rules. *Linguistic Inquiry*, 1:291–306.
- Legendre, G., Miyata, Y., and Smolensky, P. (1990). Can connectionism contribute to syntax? harmonic grammar, with an application. *CLS*, 26:237–252.
- Marcus, G. (2020). The next decade in AI: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*.
- McCarthy, J. (1998). Sympathy and phonological opacity. U. Mass., ROA.
- McCarthy, J. and Prince, A. (1993). Prosodic morphology: Constraint interaction and satisfaction. University of Massachusetts, Amherst.
- Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.
- Pater, J. (2019). Generative linguistics and neural networks at 60: foundation, friction, and fusion. *Language*, 95:41–74.
- Prince, A. and Smolensky, P. (1993). *Optimality Theory*. University of Massachusetts, Amherst and University of Colorado at Boulder.
- Rahbar, E. R. (2012). *Aspects of Persian phonology and morpho-phonology*. PhD thesis, University of Toronto.
- Smolensky, P. (2006). Harmony in linguistic cognition. *Cognitive Science*, 30:779–801.

# Selected references II

- Smolensky, P. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. ROA #1286.
- Toosarvandani, M. D. (2004). Vowel length in modern Farsi. *Journal of the Royal Asiatic Society*, 14:241–251.
- Townsend, D. J. and Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. MIT Press.
- Valiant, L. G. (2008). Knowledge infusion: In pursuit of robustness in artificial intelligence. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (2008)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik.
- Winters, T., Marra, G., Manhaeve, R., and De Raedt, L. (2021). Deepstochlog: Neural stochastic logic programming. *arXiv preprint arXiv:2106.12574*.