One (semi)ring to rule them all: Reconciling categorical and gradient models of phonotactics

LSA Session on Formal Language Theory in Morphology and Phonology

Connor Mayer

January 6th, 2024

Department of Language Science
University of California, Irvine

What is this talk about?

Question: Are phonotactic grammars categorical or gradient?

What is this talk about?

Question: Are phonotactic grammars categorical or gradient?

Answer: It depends on which semiring you use!

What is this talk about?

Two points I want to make:

What is this talk about?

Two points I want to make:

1. Gradient phonotactic models account for new data from a Turkish acceptability judgment task better than categorical models.

What is this talk about?

Two points I want to make:

1. Gradient phonotactic models account for new data from a Turkish acceptability judgment task better than categorical models.

2. This distinction turns out to be somewhat superficial if we think of models from a semiring-general perspective.

What is phonotactics?

What is phonotactics?

The legal ways in which sounds can be sequenced into words.

What is phonotactics?

The legal ways in which sounds can be sequenced into words.

This is (mostly) learned and language-specific:

What is phonotactics?

The legal ways in which sounds can be sequenced into words.

This is (mostly) learned and language-specific:

- /stik/ would be an ok English word; not a good Spanish word

What is phonotactics?

The legal ways in which sounds can be sequenced into words.

This is (mostly) learned and language-specific:

- /stik/ would be an ok English word; not a good Spanish word
- /ʧknɔw̃ɲʨ/ is a fine Polish word; not a good English word

Phonotactics is gradient

A typical source of data is to ask speakers for *acceptability judgments*:

Phonotactics is gradient

A typical source of data is to ask speakers for *acceptability judgments*:

- “on a scale of 1-7, how likely is ‘steek’ to become an English word?”
- “would ‘steek’ be a better English word than ‘chknonch’?”
- “could ‘steek’ be an English word?”

Phonotactics is gradient

A typical source of data is to ask speakers for *acceptability judgments*:

- "on a scale of 1-7, how likely is 'steek' to become an English word?"
- "would 'steek' be a better English word than 'chknonch'?"
- "could 'steek' be an English word?"

These judgments consistently display *gradience* [e.g. Chomsky and Halle, 1965, Coleman and Pierrehumbert, 1997, Scholes, 1966, Bailey and Hahn, 2001, Hayes and Wilson, 2008, Daland et al., 2011, a.o.].

What do we mean by gradience?

What do we mean by gradience?

poik

lvag

kip

What do we mean by gradience?

lvag $\ll$ poik $\ll$ kip

Where does this gradience come from?

Gradience in acceptability judgments can arise from performance factors such as misperception [e.g. Kahng and Durvasula, 2023].

Where does this gradience come from?

Gradience in acceptability judgments can arise from performance factors such as misperception [e.g. Kahng and Durvasula, 2023].

However, the gradience observed in phonotactic acceptability judgments is largely predictable from "soft" versions of the same constraints that govern other phonological processes [Hayes, 2000].

Where does this gradience come from?

Gradience in acceptability judgments can arise from performance factors such as misperception [e.g. Kahng and Durvasula, 2023].

However, the gradience observed in phonotactic acceptability judgments is largely predictable from "soft" versions of the same constraints that govern other phonological processes [Hayes, 2000].

Typical modeling approach is to use a grammar that produces a gradient output.

- Often based on statistical frequencies in the lexicon.

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

$$\text{score} : \Sigma^* \rightarrow [0, 1]$$

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

$$\text{score} : \Sigma^* \rightarrow [0, 1]$$

Such a model can represent gradient acceptability judgments:

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

$$\text{score} : \Sigma^* \rightarrow [0, 1]$$

Such a model can represent gradient acceptability judgments:

$$\text{score}(\text{lvag}) = 0.01$$

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

$$\text{score} : \Sigma^* \to [0, 1]$$

Such a model can represent gradient acceptability judgments:

$$\text{score}(\text{lvag}) = 0.01 < \text{score}(\text{poik}) = 0.2$$

Implementing a gradient phonotactic grammar

Our phonotactic grammars consist of a `score` function that assigns values to words.

$$\text{score} : \Sigma^* \rightarrow [0, 1]$$

Such a model can represent gradient acceptability judgments:

$$\text{score}(\text{lvag}) = 0.01 < \text{score}(\text{poik}) = 0.2 < \text{score}(\text{kip}) = 0.4$$

Is phonotactics categorical?

Is phonotactics categorical?

Gorman [2013] argues that we have been premature in assuming the phonotactic grammar computes gradient outputs.

Is phonotactics categorical?

Gorman [2013] argues that we have been premature in assuming the phonotactic grammar computes gradient outputs.

- **Proposal**: grammar is categorical and gradience comes from other sources.

Is phonotactics categorical?

Gorman [2013] argues that we have been premature in assuming the phonotactic grammar computes gradient outputs.

- **Proposal**: grammar is categorical and gradience comes from other sources.
- A categorical grammar labels words as either grammatical or ungrammatical

Is phonotactics categorical?

Gorman [2013] argues that we have been premature in assuming the phonotactic grammar computes gradient outputs.

- **Proposal**: grammar is categorical and gradience comes from other sources.
- A categorical grammar labels words as either grammatical or ungrammatical

In particular, he claims that **categorical models do as well as or better than gradient models** in predicting phonotactic phenomena.

Past work on categorical grammars

Categorical models have been claimed to better predict:

Past work on categorical grammars

Categorical models have been claimed to better predict:

- English onset acceptability [Gorman, 2013, Durvasula, 2020, Dai, accepted]

Past work on categorical grammars

Categorical models have been claimed to better predict:

- English onset acceptability [Gorman, 2013, Durvasula, 2020, Dai, accepted]
- Polish onset acceptability [Kostyszyn and Heinz, 2022, Dai, accepted]

Past work on categorical grammars

Categorical models have been claimed to better predict:

- English onset acceptability [Gorman, 2013, Durvasula, 2020, Dai, accepted]
- Polish onset acceptability [Kostyszyn and Heinz, 2022, Dai, accepted]
- Turkish vowel distributions [Gorman, 2013, Dai, accepted]

Past work on categorical grammars

Categorical models have been claimed to better predict:

- English onset acceptability [Gorman, 2013, Durvasula, 2020, Dai, accepted]
- Polish onset acceptability [Kostyszyn and Heinz, 2022, Dai, accepted]
- Turkish vowel distributions [Gorman, 2013, Dai, accepted]
- English medial cluster distributions [Gorman, 2013]

Limitations of previous work

Limitations of previous work

1. Use a very small number of data sets, almost all about consonant clusters

Limitations of previous work

1. Use a very small number of data sets, almost all about consonant clusters
2. Authors have different definitions of “categorical”

Limitations of previous work

1. Use a very small number of data sets, almost all about consonant clusters
2. Authors have different definitions of “categorical”
3. The gradient model used in (almost) all cases is the *UCLA Phonotactic Learner*

Limitation 2: Defining categorical

Limitation 2: Defining categorical

Some “categorical” models are in fact gradient [Durvasula, 2020, Kostyszyn and Heinz, 2022].

Limitation 2: Defining categorical

Some “categorical” models are in fact gradient [Durvasula, 2020, Kostyszyn and Heinz, 2022].

- Words receive an **integer score** corresponding to number of constraint violations

Limitation 2: Defining categorical

Some “categorical” models are in fact gradient [Durvasula, 2020, Kostyszyn and Heinz, 2022].

- Words receive an **integer score** corresponding to number of constraint violations
- “Categorical” in these models means all constraint violations are penalized equally

Limitation 2: Defining categorical

Some “categorical” models are in fact gradient [Durvasula, 2020, Kostyszyn and Heinz, 2022].

- Words receive an **integer score** corresponding to number of constraint violations
- “Categorical” in these models means all constraint violations are penalized equally

These models can represent a situation where lvag $\ll$ poik $\ll$ kip.

Limitation 2: Defining categorical

Some “categorical” models are in fact gradient [Durvasula, 2020, Kostyszyn and Heinz, 2022].

- Words receive an **integer score** corresponding to number of constraint violations
- “Categorical” in these models means all constraint violations are penalized equally

These models can represent a situation where lvag $\ll$ poik $\ll$ kip.

For the sake of time I’m going to ignore these models.

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

- Words are grammatical or not

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

- Words are grammatical or not
- I'll refer to this as a **boolean** model of phonotactics

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

- Words are grammatical or not
- I'll refer to this as a **boolean** model of phonotactics

$$\text{score} : \Sigma^* \rightarrow \{0, 1\}$$

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

- Words are grammatical or not
- I'll refer to this as a **boolean** model of phonotactics

$$\text{score} : \Sigma^* \rightarrow \{0, 1\}$$

These models *cannot* represent a situation where lvag $\ll$ poik $\ll$ kip

Limitation 2: Defining categorical

Other proposed categorical models are truly categorical [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]

- Words are grammatical or not
- I'll refer to this as a **boolean** model of phonotactics

$$\text{score} : \Sigma^* \rightarrow \{0, 1\}$$

These models *cannot* represent a situation where lvag $\ll$ poik $\ll$ kip

We'll adopt this definition of categorical.

Limitation 3: The UCLA Phonotactic Learner?

Limitation 3: The UCLA Phonotactic Learner?

The UCLA Phonotactic Learner has become the poster boy for gradient phonotactics [Hayes and Wilson, 2008].

Limitation 3: The UCLA Phonotactic Learner?

The UCLA Phonotactic Learner has become the poster boy for gradient phonotactics [Hayes and Wilson, 2008].

- But it also has to learn constraints from the data!

Limitation 3: The UCLA Phonotactic Learner?

The UCLA Phonotactic Learner has become the poster boy for gradient phonotactics [Hayes and Wilson, 2008].

- But it also has to learn constraints from the data!
- Its performance is sensitive to how it is parameterized.

Limitation 3: The UCLA Phonotactic Learner?

The UCLA Phonotactic Learner has become the poster boy for gradient phonotactics [Hayes and Wilson, 2008].

- But it also has to learn constraints from the data!
- Its performance is sensitive to how it is parameterized.
- Do categorical models outperform it because it is gradient? Because of its constraint selection process? Because it has been run with sub-optimal hyperparameters?

A simpler comparison

Let's compare the performance of two proposed categorical boolean models of Turkish vowel phonotactics against a simple probabilistic bigram model with a similar structure.

A simpler comparison

Let's compare the performance of two proposed categorical boolean models of Turkish vowel phonotactics against a simple probabilistic bigram model with a similar structure.

We'll evaluate how these models predict new experimental data from a Turkish nonce word acceptability judgment task.

A new dataset of Turkish acceptability judgments

Turkish vowels

	[−back]		**[+back]**	
	[−round]	**[+round]**	**[−round]**	**[+round]**
[+high]	i	y	ɯ	u
[−high]	e	ø	a	o

Constraints on Turkish vowels

Constraints on Turkish vowels

Backness harmony: *[αback] [$-\alpha$back]

- A vowel must agree in backness with the preceding vowel.

Constraints on Turkish vowels

BACKNESS HARMONY: *[αback] [$-\alpha$back]

- A vowel must agree in backness with the preceding vowel.

ROUNDING HARMONY: *[αround] [$-\alpha$round, $+$high].

- A high vowel must agree in roundness with the preceding vowel.

Constraints on Turkish vowels

Backness harmony: *[αback] [$-\alpha$back]

- A vowel must agree in backness with the preceding vowel.

Rounding harmony: *[αround] [$-\alpha$round, $+$high].

- A high vowel must agree in roundness with the preceding vowel.

These constraints govern suffix allomorphy, but their effect is also detectable in the lexicon and in acceptability judgment tasks [Zimmer, 1969].

Our data

The data we'll look at are acceptability judgments from a large, online study.

Our data

The data we'll look at are acceptability judgments from a large, online study.

- **Participants**: 90 native Turkish speakers (38F; mostly age 25-35) recruited on Prolific

Our data

The data we'll look at are acceptability judgments from a large, online study.

- **Participants**: 90 native Turkish speakers (38F; mostly age 25-35) recruited on Prolific
- **Task**: Wug word acceptability judgments

Stimuli

Stimuli: 596 wug words with CVCVC shape

Stimuli

Stimuli: 596 wug words with CVCVC shape

- Nine words for each unique pair of vowels (8×8 total pairs)

Stimuli

Stimuli: 596 wug words with CVCVC shape

- Nine words for each unique pair of vowels (8 × 8 total pairs)
- Probability of consonants controlled for within vowel groups

Stimuli

Stimuli: 596 wug words with CVCVC shape

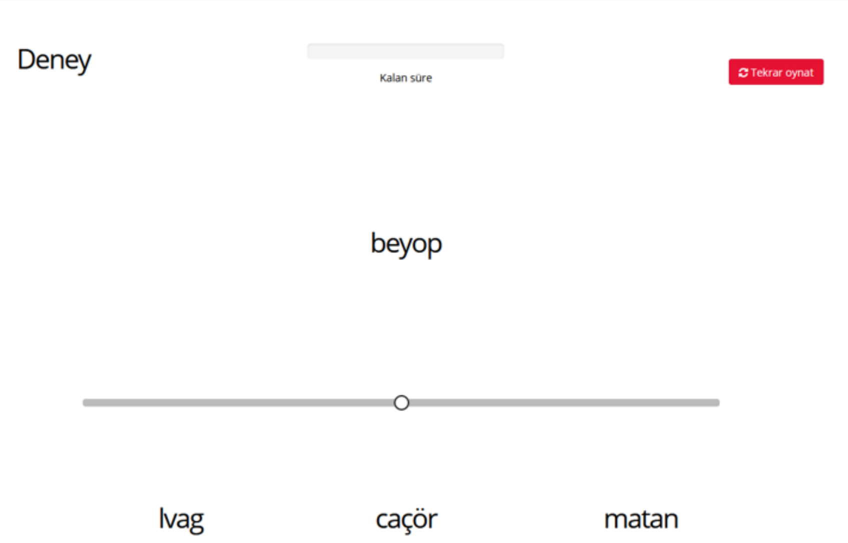
- Nine words for each unique pair of vowels (8×8 total pairs)
- Probability of consonants controlled for within vowel groups
- Synthesized to speech using Google Cloud

Stimuli

Stimuli: 596 wug words with CVCVC shape

- Nine words for each unique pair of vowels (8 × 8 total pairs)
- Probability of consonants controlled for within vowel groups
- Synthesized to speech using Google Cloud
- Words and recordings vetted by two native Turkish speakers

Experiment task



Analysis

Each participant rated 192 tokens after training and attention checks: 17,280 tokens.

Analysis

Each participant rated 192 tokens after training and attention checks: 17,280 tokens.

Responses are normalized to z-scores within participant

- Controls for differences in mean and spread between participants

Defining our models

We'll test three simple models that have similar structures:

Value type	**Constraint values**
Probability	Conditional probabilities
Boolean	Harmony [Gorman, 2013]
Boolean	Exception filtering [Dai, accepted]

General model structure

All the models are TSL-2 grammars that operate on the vowel tier

General model structure

All the models are TSL-2 grammars that operate on the vowel tier

- Informally, we **ignore consonants** and assign scores based on **vowel bigrams**

General model structure

All the models are TSL-2 grammars that operate on the vowel tier

- Informally, we **ignore consonants** and assign scores based on **vowel bigrams**

- Constraints can reference start and end symbols $\rtimes$ and $\ltimes$

Scoring bigrams

Each model type has a Δ function that assigns a value to a bigram.

Scoring bigrams

Each model type has a Δ function that assigns a value to a bigram.

Boolean model

$$\Delta_b : \Sigma^2 \to \{0, 1\}$$

Scoring bigrams

Each model type has a Δ function that assigns a value to a bigram.

Boolean model

$$\Delta_b : \Sigma^2 \rightarrow \{0, 1\}$$

Probability model

$$\Delta_p : \Sigma^2 \rightarrow [0, 1]$$

Scoring words

Scoring words

Boolean model: words are assigned 1 if they contain only legal bigrams, 0 otherwise

$$\text{bigram\_score}(x_1, \ldots, x_n) = \bigwedge_{i=1}^{n-1} \Delta_b(x_i, x_{i+1})$$

Scoring words

Boolean model: words are assigned 1 if they contain only legal bigrams, 0 otherwise

$$\text{bigram\_score}(x_1, \ldots, x_n) = \bigwedge_{i=1}^{n-1} \Delta_b(x_i, x_{i+1})$$

Probability model: words are assigned the product of the probability of each bigram.

$$\text{probability\_score}(x_1, \ldots, x_n) = \prod_{i=1}^{n-1} \Delta_p(x_i, x_{i+1})$$

An example

An example

Boolean model

$$\text{boolean\_score}([\text{oi}]) = \Delta_b(\rtimes \text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes)$$

An example

Boolean model

$$\begin{aligned}\text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1\end{aligned}$$

An example

Boolean model

$$\begin{aligned}\text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1 \\ &= 0\end{aligned}$$

An example

Boolean model

$$\begin{aligned}\text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1 \\ &= 0\end{aligned}$$

Probabilistic model

An example

Boolean model

$$\begin{aligned} \text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1 \\ &= 0 \end{aligned}$$

Probabilistic model

$$\text{probability\_score}([\text{oi}]) = \Delta_p(\rtimes\text{o}) \times \Delta_p(\text{oi}) \times \Delta_p(\text{i}\ltimes)$$

An example

Boolean model

$$\begin{aligned} \text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1 \\ &= 0 \end{aligned}$$

Probabilistic model

$$\begin{aligned} \text{probability\_score}([\text{oi}]) &= \Delta_p(\rtimes\text{o}) \times \Delta_p(\text{oi}) \times \Delta_p(\text{i}\ltimes) \\ &= 0.08 \times 0.107 \times 0.458 \end{aligned}$$

An example

Boolean model

$$\begin{aligned}\text{boolean\_score}([\text{oi}]) &= \Delta_b(\rtimes\text{o}) \wedge \Delta_b(\text{oi}) \wedge \Delta_b(\text{i}\ltimes) \\ &= 1 \wedge 0 \wedge 1 \\ &= 0\end{aligned}$$

Probabilistic model

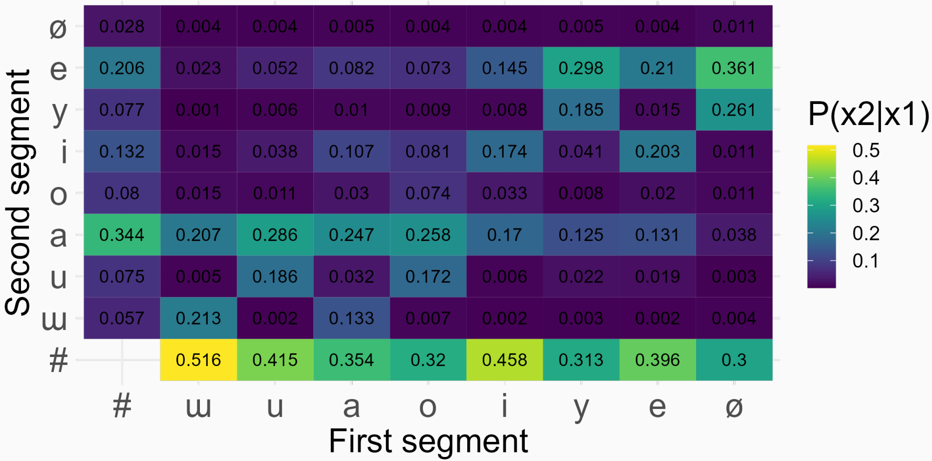
$$\begin{aligned}\text{probability\_score}([\text{oi}]) &= \Delta_p(\rtimes\text{o}) \times \Delta_p(\text{oi}) \times \Delta_p(\text{i}\ltimes) \\ &= 0.08 \times 0.107 \times 0.458 \\ &= 0.0004\end{aligned}$$

Choosing our values

How do we define Δ for each model?

Conditional probability model

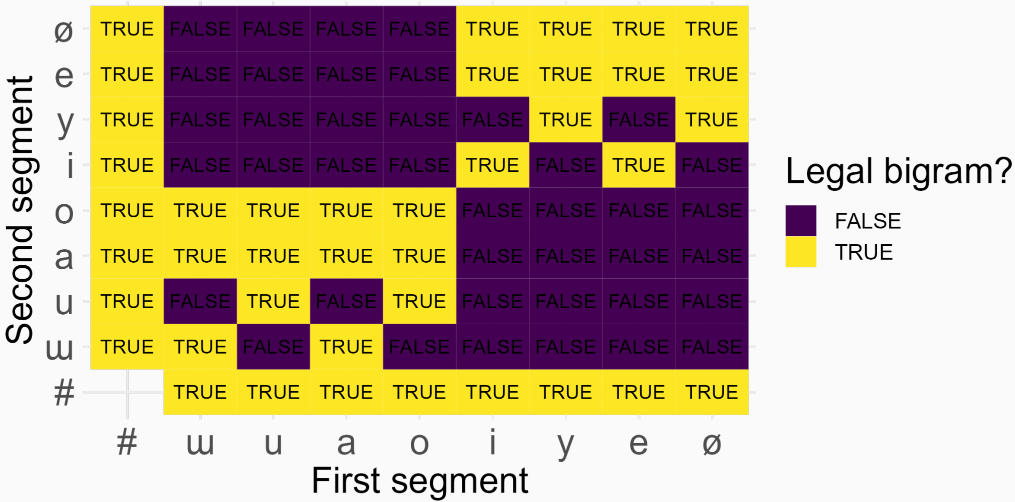
The probability model uses Laplace-smoothed conditional probabilities derived from 18,472 citation forms in the TELL database [Inkelas et al., 2000].



	#	ɯ	u	a	o	i	y	e	ø
ø	0.028	0.004	0.004	0.005	0.004	0.004	0.005	0.004	0.011
e	0.206	0.023	0.052	0.082	0.073	0.145	0.298	0.21	0.361
y	0.077	0.001	0.006	0.01	0.009	0.008	0.185	0.015	0.261
i	0.132	0.015	0.038	0.107	0.081	0.174	0.041	0.203	0.011
o	0.08	0.015	0.011	0.03	0.074	0.033	0.008	0.02	0.011
a	0.344	0.207	0.286	0.247	0.258	0.17	0.125	0.131	0.038
u	0.075	0.005	0.186	0.032	0.172	0.006	0.022	0.019	0.003
ɯ	0.057	0.213	0.002	0.133	0.007	0.002	0.003	0.002	0.004
#		0.516	0.415	0.354	0.32	0.458	0.313	0.396	0.3

Boolean harmony model [Gorman, 2013]

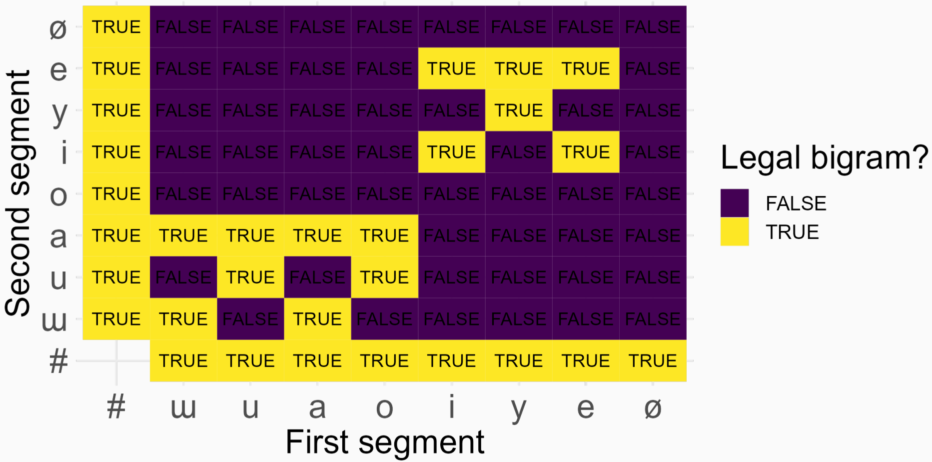
Words are grammatical if they satisfy both rounding and backness harmony.



	#	ɯ	u	a	o	i	y	e	ø
ø	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
e	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
y	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE
i	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
o	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
a	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
u	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
ɯ	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
#		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Boolean exception filtering model [Dai, accepted]

Categorical Turkish phonotactic grammar from Dai [accepted] learned via an exception filtering process.



	#	ɯ	u	a	o	i	y	e	ø
ø	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
e	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE
y	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
i	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
o	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
a	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
u	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
ɯ	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
#		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Results

Let's look at correlations between model score and mean acceptability judgment.

Results

Let's look at correlations between model score and mean acceptability judgment.

Value type	**Constraint set**	r	τ	ρ
Probability	Conditional probabilities	**0.558**	**0.375**	**0.527**
Boolean	Harmony [Gorman, 2013]	0.371	0.303	0.369
Boolean	Exception filtering [Dai, accepted]	0.360	0.286	0.348

Results

Let's look at correlations between model score and mean acceptability judgment.

Value type	**Constraint set**	r	τ	ρ
Probability	Conditional probabilities	**0.558**	**0.375**	**0.527**
Boolean	Harmony [Gorman, 2013]	0.371	0.303	0.369
Boolean	Exception filtering [Dai, accepted]	0.360	0.286	0.348

The simple probabilistic model substantially outperforms the other models

Reconciling categorical and gradient models using semirings

The reconciliation begins

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

- **Boolean**: Assigns booleans to segmental bigrams, combines them using $\wedge$.

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

- **Boolean**: Assigns booleans to segmental bigrams, combines them using $\wedge$.
- **Probabilistic**: Assigns probabilities to segmental bigrams, combines using $+$.

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

- **Boolean**: Assigns booleans to segmental bigrams, combines them using $\wedge$.
- **Probabilistic**: Assigns probabilities to segmental bigrams, combines using $+$.

But the basic structure of each model is the same:

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

- **Boolean**: Assigns booleans to segmental bigrams, combines them using $\wedge$.
- **Probabilistic**: Assigns probabilities to segmental bigrams, combines using $+$.

But the basic structure of each model is the same:

- We assign some **value** to each segmental bigram

Commonalities between boolean and probabilistic models

Probabilistic and boolean TSL-2 models differ:

- **Boolean**: Assigns booleans to segmental bigrams, combines them using $\wedge$.
- **Probabilistic**: Assigns probabilities to segmental bigrams, combines using $+$.

But the basic structure of each model is the same:

- We assign some **value** to each segmental bigram
- We **aggregate** those values to get a score for the word

Commonalities between categorical and structural models

We can abstract away from specific values/aggregators:

Commonalities between categorical and structural models

We can abstract away from specific values/aggregators:

$$\Delta\colon\ \Sigma^2 \longrightarrow \mathcal{R}$$

where $\mathcal{R}$ is some set of values

Commonalities between categorical and structural models

We can abstract away from specific values/aggregators:

$$\Delta\colon \Sigma^2 \longrightarrow \mathcal{R}$$

$$\text{score}(x_1 \ldots x_n) = \bigcirc\!\!\!\!\!\!\bigwedge_{i=1}^{n-1} \Delta(x_i, x_{i+1})$$

where $\mathcal{R}$ is some set of values and $\bigcirc\!\!\!\!\!\wedge$ is some binary operator over $\mathcal{R}$.

Other values of $\mathcal{R}$ and $\bigcirc\!\!\!\!\wedge$

We can make these simple models compute even more interesting quantities!

Other values of $\mathcal{R}$ and $\bigcirc\!\!\!\!\wedge$

We can make these simple models compute even more interesting quantities!

What does it compute?	$\mathcal{R}$	$\bigcirc\!\!\!\!\wedge$
Boolean scores [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]	$\{0, 1\}$	$\wedge$
Probabilities	$[0, 1]$	$\times$

Other values of $\mathcal{R}$ and $\owedge$

We can make these simple models compute even more interesting quantities!

What does it compute?	$\mathcal{R}$	$\owedge$
Boolean scores [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]	$\{0,1\}$	$\wedge$
Probabilities	$[0,1]$	$\times$
Integer scores [Durvasula, 2020, Kostyszyn and Heinz, 2022]	$\mathbb{N}$	$+$

Other values of $\mathcal{R}$ and $\owedge$

We can make these simple models compute even more interesting quantities!

What does it compute?	$\mathcal{R}$	$\owedge$
Boolean scores [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]	$\{0, 1\}$	$\wedge$
Probabilities	$[0, 1]$	$\times$
Integer scores [Durvasula, 2020, Kostyszyn and Heinz, 2022]	$\mathbb{N}$	$+$
Constraint violation profiles	$\mathbb{N}^k$	$+$

Other values of $\mathcal{R}$ and $\bigcirc\!\!\!\!\wedge$

We can make these simple models compute even more interesting quantities!

What does it compute?	$\mathcal{R}$	$\bigcirc\!\!\!\!\wedge$
Boolean scores [Gorman, 2013, Kostyszyn and Heinz, 2022, Dai, accepted]	$\{0,1\}$	$\wedge$
Probabilities	$[0,1]$	$\times$
Integer scores [Durvasula, 2020, Kostyszyn and Heinz, 2022]	$\mathbb{N}$	$+$
Constraint violation profiles	$\mathbb{N}^k$	$+$
Left SL-2 string transduction	Σ^*	$+$

What’s going on here?

What's going on here?

This definition of our TSL-2 models is in **semiring-general terms**

$$\Delta\colon\ \Sigma^2 \longrightarrow \mathcal{R}$$

$$\text{score}(x_1 \dots x_n) = \bigcirc\!\!\!\!\!\!\bigwedge_{i=1}^{n-1} \Delta(x_i, x_{i+1})$$

What's going on here?

This definition of our TSL-2 models is in **semiring-general terms**

$$\Delta\colon\ \Sigma^2 \longrightarrow \mathcal{R}$$

$$\text{score}(x_1 \ldots x_n) = \bigotimes_{i=1}^{n-1} \Delta(x_i, x_{i+1})$$

We can parameterize our model with different semirings that provide implementations of $\mathcal{R}$ and $\bigotimes$.

What's a semiring?

A semiring is an algebraic structure.

What's a semiring?

A semiring is an algebraic structure.

Monoid: a set $\mathcal{R}$ closed under a binary relation $\bigcirc\!\!\!\!\wedge$ such that:

What’s a semiring?

A semiring is an algebraic structure.

Monoid: a set $\mathcal{R}$ closed under a binary relation $\owedge$ such that:

- $\owedge$ is associative

What's a semiring?

A semiring is an algebraic structure.

Monoid: a set $\mathcal{R}$ closed under a binary relation $\bigcirc\!\!\!\!\wedge$ such that:

- $\bigcirc\!\!\!\!\wedge$ is associative
- There's an identity element $\top$ in $\mathcal{R}$ such that $a \bigcirc\!\!\!\!\wedge \top = \top \bigcirc\!\!\!\!\wedge a = a$

Semirings

A **semiring** consists of a pair of monoids

Semirings

A **semiring** consists of a pair of monoids

- $(\mathcal{R}, \bigcircle{\wedge})$ with identity element $\top$

Semirings

A **semiring** consists of a pair of monoids

- $(\mathcal{R}, \bigcirc\!\!\!\!\!\wedge\,)$ with identity element $\top$
- $(\mathcal{R}, \bigcirc\!\!\!\!\!\vee\,)$ with identity element $\bot$

Semirings

A **semiring** consists of a pair of monoids

- $(\mathcal{R}, \bigwedge\!\!\!\!\bigcirc)$ with identity element $\top$
- $(\mathcal{R}, \bigvee\!\!\!\!\bigcirc)$ with identity element $\bot$

such that:

- $\bigwedge\!\!\!\!\bigcirc$ distributes over $\bigvee\!\!\!\!\bigcirc$

Semirings

A **semiring** consists of a pair of monoids

- $(\mathcal{R}, \bigowedge)$ with identity element $\top$
- $(\mathcal{R}, \bigovee)$ with identity element $\bot$

such that:

- $\bigowedge$ distributes over $\bigovee$
- $x \bigowedge \bot = \bot \bigowedge x = \bot$

Why are semirings interesting?

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

- In terms of $\mathcal{R}$, $\owedge$, $\ovee$ rather than specific values and operators

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

- In terms of $\mathcal{R}$, $\owedge$, $\ovee$ rather than specific values and operators
- (TSL-2 models don't use $\ovee$ but it's important for more complex models)

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

- In terms of $\mathcal{R}$, $\owedge$, $\ovee$ rather than specific values and operators
- (TSL-2 models don't use $\ovee$ but it's important for more complex models)

Different semirings allow the same underlying model to compute different quantities.

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

- In terms of $\mathcal{R}$, $\owedge$, $\ovee$ rather than specific values and operators
- (TSL-2 models don't use $\ovee$ but it's important for more complex models)

Different semirings allow the same underlying model to compute different quantities.

- Unifies superficially different models [Goodman, 1999].

Why are semirings interesting?

The models we work with in FLT (TSL, FSA, CFG, etc.) can be expressed in semiring-general terms.

- In terms of $\mathcal{R}$, $\owedge$, $\ovee$ rather than specific values and operators
- (TSL-2 models don't use $\ovee$ but it's important for more complex models)

Different semirings allow the same underlying model to compute different quantities.

- Unifies superficially different models [Goodman, 1999].

We can separate the structure of the model from the values it computes.

Why is this useful for us as phonologists?

Why is this useful for us as phonologists?

Semirings allow us to relate the grammar to different domains or contexts

Why is this useful for us as phonologists?

Semirings allow us to relate the grammar to different domains or contexts

Giorgolo, G., & Asudeh, A. (2014) One semiring to rule them all. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Québec City: Cognitive Science Society, 208–26.

Why is this useful for us as phonologists?

Semirings allow us to relate the grammar to different domains or contexts

- Giorgolo and Asudeh [2014] apply different semirings to the same underlying semantic model to capture differences in heuristic vs. mathematical reasoning.

Giorgolo, G., & Asudeh, A. (2014) One semiring to rule them all. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Québec City: Cognitive Science Society, 208–26.

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

- Harmony is essentially categorical when determining suffix allomorphy

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

- Harmony is essentially categorical when determining suffix allomorphy

‘cat-PL‘

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

- Harmony is essentially categorical when determining suffix allomorphy

'cat-PL'	kedi-ler ✓

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

- Harmony is essentially categorical when determining suffix allomorphy

‘cat-PL‘ kedi-ler ✓ kedi-lar ✗

- Harmony is a gradient preference when determining word acceptability

Connecting the grammar to different domains

There's perhaps an analogy to be made to Turkish.

- Harmony is essentially categorical when determining suffix allomorphy

  ‘cat-PL‘ kedi-ler ✓ kedi-lar ✗

- Harmony is a gradient preference when determining word acceptability
- **But both sensitive to the same configurations!**

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

- are sensitive only to bigram constraints

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

- are sensitive only to bigram constraints
- use segmental representations

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

- are sensitive only to bigram constraints
- use segmental representations
- operate on the vowel tier

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

- are sensitive only to bigram constraints
- use segmental representations
- operate on the vowel tier

These are segmental TSL-2 grammars, regardless of the values they assign.

Connecting the grammar to different domains

Regardless of semiring, both the categorical and probabilistic grammars we saw here

- are sensitive only to bigram constraints
- use segmental representations
- operate on the vowel tier

These are segmental TSL-2 grammars, regardless of the values they assign.

The same applies to other representations or grammars.

Closing remarks

Durvasula [2020] closes with a plea to abandon gradience and adopt categorical grammars so we can

Closing remarks

Durvasula [2020] closes with a plea to abandon gradience and adopt categorical grammars so we can

- “focus on what’s a possible constraint or rule”; and

Closing remarks

Durvasula [2020] closes with a plea to abandon gradience and adopt categorical grammars so we can

- "focus on what's a possible constraint or rule"; and
- "commit to a specific set of representations"

We can have our phonotactic cake and eat it too

This is a false dichotomy.

We can have our phonotactic cake and eat it too

This is a false dichotomy.

- Constraints and representations in the grammar can be studied independently of the values the grammar assigns.

We can have our phonotactic cake and eat it too

This is a false dichotomy.

- Constraints and representations in the grammar can be studied independently of the values the grammar assigns.

- Insight into the structure of the grammar can come from both gradient and categorical analyses!

We can have our phonotactic cake and eat it too

This is a false dichotomy.

- Constraints and representations in the grammar can be studied independently of the values the grammar assigns.
- Insight into the structure of the grammar can come from both gradient and categorical analyses!
- This flexibility allows our models to engage with a broader range of empirical phenomena.

Thank you!

Thanks to Huteng Dai, Jon Rawski, Megha Sundara, and Richard Futrell for many interesting discussions, and to my Turkish consultants Cem Babalik and Defne Bilhan.

References i

References

Noam Chomsky and Morris Halle. Some controversial questions in phonological theory. *Journal of Linguistics*, 1(2):97–138, 1965.

John Coleman and Janet Pierrehumbert. Stochastic phonological grammars and acceptability. In John Coleman, editor, *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology*, pages 49–56. Association for Computational Linguistics, Somerset, NJ, 1997.

Robert Scholes. *Phonotactic grammaticality*. Mouton, The Hague, 1966.

Todd M. Bailey and Ulrike Hahn. Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language*, 44: 568–591, 2001.

Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440, 2008.

Robert Daland, Bruce Hayes, James White, Marc Garellek, Andreas Davis, and Ingrid Normann. Explaining sonority projection effects. *Phonology*, 28:197–234, 2011.

Jimin Kahng and Karthik Durvasula. Can you judge what you don't hear? perception as a source of gradient wordlikeness judgments. *Glossa*, 8(1), 2023. doi: \url{https://doi.org/10.16995/glossa.9333}.

Bruce Hayes. *Gradient well-formedness in Optimality Theory*, pages 88–120. Oxford University Press, 2000.

Kyle Gorman. *Generative phonotactics*. PhD thesis, University of Pennsylvania, 2013.

Karthik Durvasula. O gradience, whence do you come?, 2020. Keynote presentation at the 2020 Annual Meeting on Phonology.

References ii

Huteng Dai. An exception-filtering approach to phonotactic learning, accepted.

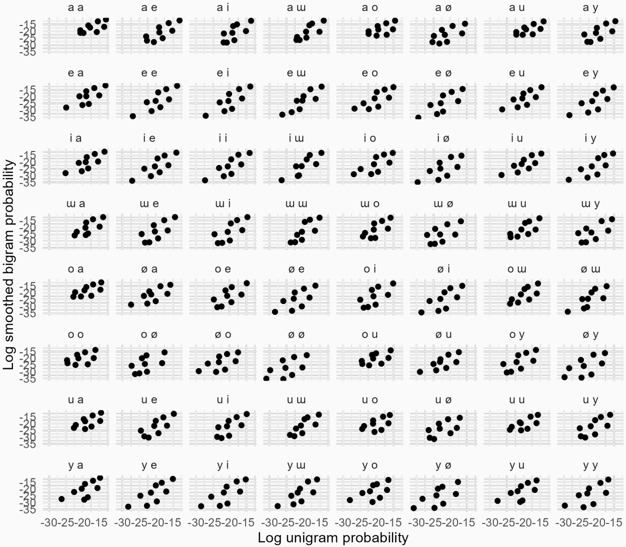
Kalina Kostyszyn and Jeffrey Heinz. Categorical account of gradient acceptability of word-initial Polish onsets. In *Proceedings of AMP 2021*. 2022.

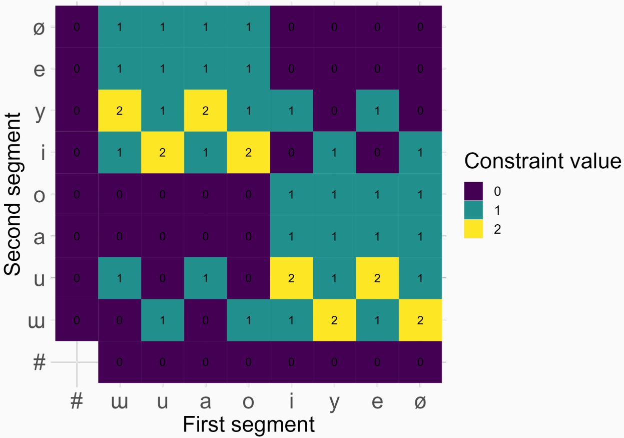
K.E. Zimmer. Psychological correlates of some Turkish morpheme structure conditions. *Language*, pages 309–321, 1969.

S. Inkelas, A. Küntay, C.O. Orgun, and R. Sprouse. Turkish electronic living lexicon (TELL): A lexical database. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*. European Language Resources Association (ELRA), Athens, Greece, 2000.

Josh Goodman. Semiring parsing. *Computational Linguistics*, 25(4):573–605, 1999.

G. Giorgolo and A. Asudeh. One semiring to rule them all. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 208–226. Cognitive Science Society, Québec City, 2014.

Stimuli structure

Cost semiring



	#	ɯ	u	a	o	i	y	e	ø
ø	0	1	1	1	1	0	0	0	0
e	0	1	1	1	1	0	0	0	0
y	0	2	1	2	1	1	0	1	0
i	0	1	2	1	2	0	1	0	1
o	0	0	0	0	0	1	1	1	1
a	0	0	0	0	0	1	1	1	1
u	0	1	0	1	0	2	1	2	1
ɯ	0	0	1	0	1	1	2	1	2
#		0	0	0	0	0	0	0	0

Results

Value type	**Constraint set**	r	τ	ρ
Probability	Conditional probabilities	**0.558**	**0.375**	**0.527**
Boolean	Cost [Durvasula, 2020, Kostyszyn and Heinz, 2022]	-0.379	-0.305	-0.386
Boolean	Harmony [Gorman, 2013]	0.371	0.303	0.369
Boolean	Exception filtering [Dai, accepted]	0.360	0.286	0.348