Morpheme structure phonotactics: a categorical model for morpho-phonological productivity in Russian vowel-zero alternations

Daniar Kasenov
New York University
dk4981@nyu.edu

Abstract

Nonce word studies motivate a notion of gradient *similarity* between nonce words and real words. In morpho-phonological research, similarity is often taken as to be a relationship between a nonce word and the list of morphemes / words that undergo a given morpho-phonological alternation (Albright and Hayes 2003; Becker et al. 2011 i.a.). This paper challenges this view on the basis of nonce word data on Russian vowel–zero alternations (Gouskova and Becker 2013; Becker and Gouskova 2016). I propose a model where morpho-phonological similarity is a relationship between the available underlying representations and the underlying representation the nonce item must have in order to undergo the alternation. The implementation of the proposed model matches—and in some comparisons exceeds—the performance of Becker and Gouskova's (2016) MaxEnt-model. This study thus presents a linking hypothesis between nonce word studies and approaches that mark segments themselves as undergoing certain restricted alternations.

1 Introduction

Some phonological alternations only apply to certain morphemes (they are *morpho*-phonological: see Zimmermann 2026 for an overview). Such alternations show gradient productivity: their rate of application to nonce words may lie anywhere between 0 and 1 depending on the properties of the nonce word (Albright and Hayes 2003; Hayes et al. 2009; Becker et al. 2011; Becker and Gouskova 2016 i.a.). Consider Russian vowel-zero alternations like *mox* ∼ *mx-a* 'moss', compared to *nos* ∼ *nos-a* 'nose' where the alternation does not apply (Scheer 2011; Gouskova 2012 i.a.). A Russian speaker is more likely to accept a nonce word alternation like *posot*∼*post-a* than an alternation like *posort*∼*posrt-a*, due to syllable structure related restrictions (Gouskova and Becker 2013).

In line with most work in phonotactics, these findings have been interpreted as evidence that linguistic knowledge provides a inherently gradient notion of *similarity* between nonce words and real words, often modelled using a weighted constraints grammar (Hayes and Wilson 2008; Wilson and Gallagher 2018; Mayer and Nelson 2020; Dai et al. 2023; Mayer 2025). For example, the proposal of Becker and Gouskova (2016) relies on Hayes and Wilson's MaxEnt-based phonotactic model.

For morpho-phonological purposes, similarity is often defined as how similar the nonce word is to real words that undergo a morpho-phonological process (so-called sublexicons: Becker et al. 2011; Gouskova and Becker 2013). On this basis, nonce word studies are taken to support theories that mark whole morphemes, rather than particular segments as exceptional targets or triggers of morpho-phonological processes (Gouskova 2012).

This paper challenges these views on the basis of Russian vowel-zero alternations. I examine whether the experimental results of Becker and Gouskova (2016) require a weighted constraints-based phonotactic model or morpheme marking of exceptionality. First, I implement Becker and Gouskova's proposal using a phonotactic model where constraints are equally weighted—to probe whether constraint weights do any real work in predicting the results (see Gorman 2013; Kostyszyn and Heinz 2021; Durvasula 2026 for similar skepticism vis-à-vis phonotactic studies).

Second, I entertain a model for morpho-phonological productivity that relies on evaluating possible underlying representations of a nonce word relative to the underlying representations that already exist in the lexicon. I dub the model *morpheme structure phonotactics*: the speakers are able to evaluate possible URs for a nonce word based on their phonotactic properties relative to real URs, echoing morpheme structure constraints used in phonological theory (Halle 1959; Booij 2011).

Proceedings of the Society for Computation in Linguistics (SCiL) 2026, pages 147–159
July 3-4, 2026 ©2026 Association for Computational Linguistics

As noted by an anonymous reviewer, the proposal is similar to existing work on Dutch voicing alternations: Ernestus and Baayen (2003) find that surrounding segments are a major predictor of whether a nonce word-final obstruent will alternate. In their own words, "speakers are more likely to choose that phoneme as the underlying representation when there are more phonologically/phonetically similar morphemes in the lexicon sharing that phoneme" (ibid.: 29)—this paper extends such attitude to cases of abstract phonemes (since segment-marking of exceptionality often relies on contrasts that are neutralized on the surface).

All three models end up with a similar fit to the nonce word judgement data. On the other hand, theoretical morpho-phonological work has established a number of cases where segments themselves must be the targets of morpho-phonological processes (see Martínez 2008 on Modern Hebrew spirantization; Rubach 2013, 2016 on Polish vowel-zero alternations; Round 2017 on Yidiny final vowel deletion *i.a.*). Thus, while other sources of evidence favor segment-based view of morpho-phonology, nonce word studies do not provide evidence against it—the resulting picture is then favorable for the segment-based view.

The paper is structured as follows. Section 2 introduces two competing models for morpho-phonological productivity that rely on phonotactics. Section 3 introduces the issues surrounding gradience in phonotactic modelling. Section 4 presents the comparison of the models on the dataset of Becker and Gouskova (2016). Section 5 concludes.

2 Morpho-phonological models

Morpho-phonological alternations such as the Russian vowel–zero pattern raise the issue of encoding: since the alternations are exceptional, the lexicon must mark whether a morpheme undergoes the alternation or not. The opposing positions are morpheme- and segment-marking of exceptionality (see Gouskova 2012 for an overview of the debate). Coming back to the example from earlier, the question is whether a feature that conditions vowel deletion, call it F, is a property of the whole morpheme: *mox∼mx-a* 'moss' is +F whereas *nos∼nos-a* isn't, or a property of the vowel itself: the vowel *o* of *mox∼mx-a* 'moss' is underlyingly +F whereas the vowel *o* of *nos∼nos-a* isn't. Note that the latter view is often committed to positing "abstract" contrasts that do not surface phonetically.

Nonce word productivity results has been interpreted as evidence in favor of morpheme-marking of exceptionality (Becker et al. 2011; Becker and Gouskova 2016 *i.a.*). The argument goes: phonotactic tendencies that only hold of items that undergo an exceptional alternation predict whether a nonce word can undergo that alternation. Therefore, the full segmental composition of the morphemes that undergo an alternation must be available for the speaker to analogize over in a nonce word task.

In this section, I outline two models for morpho-phonological productivity. The first model is the established sublexical model (Becker et al. 2011; Allen and Becker 2015; Becker and Gouskova 2016), and relies on morpheme-marking of exceptionality. The other model is the *morpheme structure phonotactics* model, loosely inspired by Ernestus and Baayen (2003): phonotactic tendencies in morpho-phonological productivity are due to distribution of underlying segments in the lexicon. The question, then, becomes a question of locality: whether the immediate surrounding environment of the segment targeted by an exceptional alternation suffices to capture the phonotactic tendencies. I elaborate on the problem in Section 2.3.

2.1 A morpheme exceptionality model

Many theorists reify the observation about morphemes (not) undergoing an alternation: the phonological alternation is posited to apply only if the vowel is a part of a morpheme marked with a diacritic feature z. This idea makes it part of linguistic knowledge whether any given morpheme undergoes the alternation (Pater 2008; Gouskova 2012). E.g., the morpheme *konets* 'end.NOM.SG' that surfaces without the second vowel in *konts-a* 'end-GEN.SG' will bear z while the non-alternating *master* 'master.NOM.SG' ∼ *master-a* 'master-GEN.SG' will not. The alternation itself is driven by a markedness constraint like *MID$_z$ which penalizes mid vowels /e,o/ in realizations of morphemes marked by z, favoring vowel deletion (Gouskova 2012).

Such analysis makes available a list of morphemes that are marked with z, a sublexicon. Gouskova and Becker (2013) show that the sublexicon with alternating vowels has phonotactic tendencies that do not hold of the lexicon as a whole. For example, words that end with a CC sequence like *most* 'bridge' are 17% of masculine nouns in Russian, **no** masculine nouns with an alternating vowel end with a CC sequence. Nonce word tasks show that speakers are aware of this fact (ibid.).

	*CC]	*$\textsc{Mid}_z$	Max
a. /nos-a/			
☞ [nos-a]			
[ns-a]			*!
b. /mox$_z$-a/			
☞ [mx-a]			*
[mox-a]		*!	
c. /mox$_z$/			
☞ [mox]		*	
[mx]	*!		*

Figure 1: The basic logic of Gouskova's analysis: $\textsc{Mid}_z$ over Max forces deletion which is blocked by high ranking *CC] constraint.

The sublexical morpho-phonological model (Becker et al. 2011; Allen and Becker 2015; Becker and Gouskova 2016) makes use of the idea that the speaker's knowledge of the alternation triggered by the feature z involves phonotactic generalizations over the sublexicon of items that bear the feature z. Nonce word task is then a matter of assigning the feature z to a nonce word—thus, evaluating whether the it conforms to existing the generalizations about z-bearing items, i.e. how similar the nonce word is to the list of z-bearing words.

$$p(nom \sim gen_\emptyset) \propto s(nom \mid L_z) \times s(gen_\emptyset \mid L)$$

How to read the proportion: $p(nom \sim gen_\emptyset)$, the probability of accepting a form of the nonce item with vowel deletion, is proportional to the product of $s(gen_\emptyset \mid L)$, the measure of phonotactic similarity of the form with vowel deletion $gen_\emptyset$ to the whole lexicon L, and $s(nom \mid L_z)$, the measure of phonotactic similarity of the base form nom to the sublexicon of forms bearing the feature z, L_z—in other words, whether the derived form conforms to the phonotactics of the whole language and the base conforms to the phonotactics of the sublexicon.

To generalize, $p(B \sim D)$, the probability of the alternation $B \sim D$ between the base B and the derived form D, is proportional to: $s(D \mid L)$, the measure of phonotactic similarity of D to the lexicon L, $s(B \mid L_z)$, the measure of phonotactic similarity of B to a sublexicon L_z, and $p(B \mapsto_z D)$ the probability of B corresponding to D given the diacritic z.

$$p(B \sim D) \propto s(B \mid L_z) \times s(D \mid L) \times p(B \mapsto_z D)$$

Like Becker and Gouskova (2016), I take the grammar to be deterministic: $p(B \mapsto_z D) \in \{0, 1\}$. However, the model is fully compatible with a probabilistic phonological grammar.

2.2 A segment exceptionality model

For Russian vowel-zero alternation, the alternative to the morpheme exceptionality view is a distinction between two types of underlying vowels, which is neutralized on the surface: either a feature distinction (Lightner 1965; Halle and Matushansky 2006) or a difference in autosegmental structure (Kenstowicz and Rubach 1987; Scheer 2019 i.a.). To be non-committal, I assume the distinction between vowels to be encoded by the feature F.

The usual analysis holds that a vowel marked with F only surfaces when preceding a syllable with a vowel marked with F (Scheer 2011). The form *mox* is $mo_F xo_F$ underlyingly and the form *mx-a* is $mo_F xa$. Figure 2 shows the application of rules in (1).

(1) a. $V_F \rightarrow V$ /___C_0V_F

b. $V_F \rightarrow \emptyset$

$mo_F x$-o_F	$mo_F x$-a	nos-o_F	nos-a	UR
$mox o_F$	$mo_F xa$	$noso_F$	nosa	Rule (1a)
mox	mxa	nos	nosa	Rule (1b)

Figure 2: Derivation of the alternation.

Under the segment exceptionality view, extending an alternation to a nonce word is a matter of assigning an appropriate UR: i.e., accepting the nonce UR with a F-marked vowel (see Nelson 2024:61 for a similar statement). Gradient nonce word results will be a consequence of gradient similarity of a nonce UR to existing URs.

I propose that similarity of a nonce UR is determined by the speakers' knowledge of possible sequences of underlying segments in their lexicon: call it morpheme structure phonotactics, a generalization of morpheme structure constraints (see the overview in Booij 2011). One is more likely to accept a nonce item alternation if the required underlying form conforms to the distribution of underlying segments in their lexicon (see also Hypothesis 4 of Ernestus and Baayen 2003).

$$p(nom \sim gen_\emptyset) \propto s(nom_F \mid U) s(gen_\emptyset \mid L)$$

How to read the proportion: the probability $p(nom \sim gen_\emptyset)$ of accepting a derived form with vowel deletion is proportional to the product of $s(gen_\emptyset \mid L)$, the measure of phonotactic similarity of the derived form $gen_\emptyset$ to the lexicon L and $s(nom_F \mid U)$, the measure of phonotactic similarity of the UR nom_F to the lexicon of URs U.

To generalize, $p(B \sim D)$, the probability of the alternation $B \sim D$ between the base B and the derived form D, is proportional to: $s(D \mid L)$, the measure of phonotactic similarity of D to the lexicon L, $s(B_u \mid U)$, the measure of phonotactic similarity of the UR B_u to the lexicon of URs U, and $p(B_u \mapsto D)$, the probability of the $B_u \mapsto D$ mapping.

$$p(B \sim D) \propto s(B_u \mid U) \times s(D \mid L) \times p(B_u \mapsto D)$$

The assumption, again, is that the grammar has a single output: $p(B_U \mapsto D) \in \{0, 1\}$. However, there is no principled conflict with a probabilistic phonological grammar.

2.3 How either model captures tendencies

Let us consider how either model accounts for the phonotactic tendencies in alternating morphemes, such as the *CC# ban (Section 2.1; Gouskova and Becker 2013). To repeat: even though 17% of Russian masculine nouns end in a consonant cluster, no masculine noun undergoing the vowel-zero alternation does.

Any model based on morpheme exceptionality has no trouble capturing the generalization: if the sublexicon of alternating words has no CC-final ones, that's a strong phonotactic restriction which a nonce item is unlikely to overcome (as results of Gouskova and Becker 2013 show).

Things seem harder for a segment exceptionality model: assuming that only 1-, 2-, and 3-grams matter for phonotactics (as held by Hayes and Wilson 2008; Wilson and Gallagher 2018; Gouskova and Gallagher 2020 i.a.), the ban on a CC# sequences cannot be directly expressed with a reference to the alternating vowel. However, not all phonotactic restrictions have to be captured in a direct way.

An independent fact is that only the last vowels of morphemes may alternate (Gouskova 2012). Therefore, $\mathrm{V_F}$CC can only occur word-finally, in a $\mathrm{V_F}$CC# environment. A ban on *CC# items that undergo the alternation falls out of a ban on an underlying $\mathrm{V_F}$CC sequences together with the generalization that $\mathrm{V_F}$ must be last vowel in the morpheme (which is a consequence of alternation-driven learning: a $\mathrm{V_F}$ that isn't the last vowel of the morpheme either always surfaces as V or zero; Scheer 2019:212 makes a similar point in his discussion of Gouskova and Becker's findings).

To capture the tendencies, both model rely on an independent measure of phonotactic similarity. The next section discusses two such models.

3 The phonotactic models

There is no shortage of phonotactic models in the literature (see Mayer 2025 for a recent discussion). For the purposes of this paper, I contrast two: the gradient model of Hayes and Wilson (2008) and the categorical model of Durvasula (2026). See Durvasula's paper for a thorough discussion of the role of gradience in phonotactic modelling.

A phonotactic model is characterized by two core properties. The first is the *hypothesis space*: what kinds of generalizations can be extracted from the lexicon. Hayes and Wilson's (2008) model is able to formulate constraints on n-grams of bundles of features. For example, $*\langle \begin{bmatrix} +son \\ -lab \end{bmatrix}, \begin{bmatrix} -son \\ -lab \end{bmatrix} \rangle$ which penalizes non-labial sonorant-obstruent sequences (*nk, nd,...) is a possible constraint. On the other hand, the model of Durvasula (2026) restricts the hypothesis space to segmental n-grams and n-grams of valued features: so, $*\langle [+son], [-son] \rangle$ is a possible constraint, but $*\langle \begin{bmatrix} +son \\ -syll \end{bmatrix}, \begin{bmatrix} +son \\ +syll \end{bmatrix} \rangle$ is not, since it refers to bundles of valued features. It is possible to restrict the hypothesis space even further and only entertain segment-based n-grams. Appendix A discusses two such models.

The second core property is *gradience* of evaluation. In the model of Hayes and Wilson (2008), constraints (illicit n-grams) are unequal in their contribution to the resulting score. The harmony score is the number of times the word violates a constraint (C_n) times its weight (w_n, summed over N constraints. The constraints and their weights are determined using a MaxEnt learning mechanism, described in ibid. Hence, I refer to their model as a MaxEnt model.

$$h(x) = \Sigma_{n=1}^{N} w_n C_n(x)$$

The model of Durvasula (2026), however, treats all constraints as equally weighted. So the output of the model is the count of occurrences of a banned n-gram in the word $C_n(x)$, summed over all banned n-grams. Therefore, the weights are all equal whereas the constraints themselves are simply those in the hypothesis space which are absent from the corpus. Hence, I refer to his model as a segmental-featural cost (SF cost) model. Appendix B discusses a modification of Durvasula's model without segmental n-grams.

$$m(x) = \Sigma_{n=1}^{N} C_n(x)$$

In this work, I report the results of the segmental-featural cost models over 2- and 3-grams (SF 2,3-cost). The score of the segmental-featural cost models is transformed into probabilities of accepting / rejecting any given nonce item by taking $\frac{1}{m(x)+1}$ ($= e^{-ln(x+1)}$; a phonotactically ideal nonce item will always be accepted by participants, and so on). A probability estimate for the MaxEnt model is taken from Becker and Gouskova (2016).

Before we turn to the evaluation of the models, I wish to emphasize that the categorical / gradient divide concerns degrees of sequence illicitness (weighting). Any linking hypothesis between a phonotactic score and a nonce word task is a way to transform the score into a probability of accepting a given nonce item—so the fact that both a MaxEnt-based model and cost-based models ultimately return a value in the range from 0 to 1 should not obscure the different assumptions about weighting, and, thus, the relevance of the findings to the gradience debate in phonotactic modelling.

4 Model evaluation

This section compares the sublexical MaxEnt model of Becker and Gouskova (2016) to a morpheme exceptionality based model (the sublexical model) using an SF cost phonotactic model (to see whether the sublexical model requires any strong committments to a phonotactic model) and to a segment exceptionality based model using an SF cost phonotactic model (to see whether the morpheme structure phonotactics proposal is viable). I do not report a morpheme structure phonotactics model based on MaxEnt phonotactics due to its lack of theoretical interest: a proponent of MaxEnt grammars is unlikely to be interested in segment exceptionality and vice versa.

The corpus materials, experimental data, feature decompositions, and simulations for the MaxEnt sublexical model are taken from Becker and Gouskova and are available at https://becker.phonologist.org/projects/yers/. The materials for my own simulations are available at https://github.com/antidanyar/msp.

I first present quantitative evaluation of the models (Sections 4.3-4.4), which slightly favors the morpheme structure phonotactics model, and then discuss how the model captures the generalizations about alternating morphemes covered by Gouskova and Becker (2013), explaining its quantitative success (Section 4.5).

4.1 The corpus data & features

The Russian lexicon is approximated by the list of 2nd declension masculine nominals in their NOM.SG and GEN.SG forms, with annotation regarding vowel-zero alternations (adapted from Zaliznjak's 1977 dictionary of Russian). For example, the data for the word *šov* 'seam' involve the forms *šov* and *šv-a* (NOM.SG and GEN.SG) with an indication that it is a word that undergoes the vowel-zero alternation. An important caveat is that some words in the training data are multi-morphemic and contain derivational affixes: for example, the word *ščen-ok* ∼ *ščen-k-a* 'puppy' has the diminutive affix *-(o)k* which undergoes the alternation. As far as I am aware, this simplification does not jeopardize any claims made in the paper—I leave replication on a more realistic lexicon of URs for future work.

To get the valued feature n-grams required for the segment-featural cost models, I have used the feature decomposition employed by Becker and Gouskova (2016). To compute the phonotactic scores over URs, conservative underlying representations have been assumed, which correspond to the dictionary entries of NOM.SG with one change: vowels that undergo the vowel-zero alternation are marked as o_F and e_F to make them distinct from vowels that do not.

For the segmental-featural cost models, I have supplemented the feature decomposition with a $\pm F$ feature: o_F and e_F are $+F$ (since only /o,e/ alternate) and all other segments are $-F$. That choice did not end up being substantive: see Table 16 in Appendix for the feature matrix employed in the simulations reported here & Appendix B for a discussion of results with alternative feature sets.

4.2 The experimental data

Participants were shown a nonce item and asked to rate it on a 1-5 scale. They were then given two suffixed forms, with deletion and without, and were asked if they are acceptable: see Figure 3 for an example trial. The nonce items varied in the quality of the alternating vowel (*o* or *e*) number of syllables (monosyllabic vs. disyllabic where disyllabics have a CV first syllable), complexity of onsets/codas (CVC, CCVC, CVCC) and the sonority sequence of the variant without vowel (TR, TRT, and so on).

Becker and Gouskova (2016) note that the reported wellformedness of the base form didn't fully determine the acceptability of the derived forms, I employ those estimates as baseline model

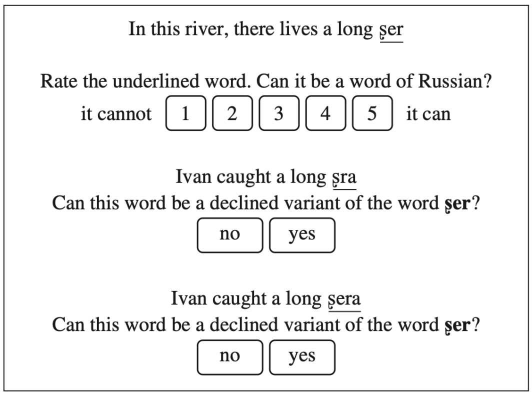

Figure 3: Translation of an example trial showing a nominative base and two genitive derivatives. Reproduced from Becker and Gouskova (2016), p.398.

(by taking $\frac{1}{6-x}$)–if anything, a successful morphophonological model must outperform the effect of the phonotactic wellformedness of the base reported by the participant.

4.3 Evaluation I: aggregates over types

Following Becker and Gouskova (2016), I evaluate the models based on the average score given to each syllabic type (their design resulted in most items only being evaluated by a single participant, see Section 3.3 of ibid.). Only aggregation by syllabic types makes it possible for correlation scores to be informative, given the binary choice task.

Table 1 presents three correlation metrics with bootstrapped 95% confidence intervals (see Albright 2009 and Durvasula 2026 for some discussion of the use of correlation metrics for evaluation of phonotactic models).

Model	Pearson r [95% CI]	Spearman ρ [95% CI]	Kendall τ_b [95% CI]
MaxEnt (morpheme)	**.893** [.837, .939]	.854 [.746, .915]	.673 [.566, .769]
SF 2,3-cost (segment)	.844 [.778, .925]	**.858** [.729, .928]	**.683** [.556, .796]
Baseline	.738 [.640, .825]	.655 [.480, .758]	.432 [.293, .560]
SF 2,3-cost (morpheme)	.713 [.605, .828]	.804 [.663, .885]	.607 [.478, .727]
SF 2-cost (morpheme)	.654 [.504, .777]	.609 [.390, .754]	.411 [.240, .561]
SF 2-cost (segment)	.552 [.344, .718]	.519 [.253, .716]	.373 [.184, .529]

Table 1: Syllabic type-level correlations between model predictions and empirical acceptability estimates. Brackets indicate 95% confidence intervals obtained by bootstrap resampling ($n = 5000$).

The aggregate correlation values suggest that the two best performing models are the sublexical MaxEnt model of Becker and Gouskova (2016) and the morpheme structure phonotactics model (SF-cost segment) over 2- and 3-grams. Based on the 95% CIs, these models are the only ones that significantly outperform the baseline.

However, no choice between the top two models is possible on the basis of the correlation metrics and their CIs, suggesting that another metric is needed. Additionally, one might be uncomfortable with evaluating model fit against averages over syllabic types—I now present a metric computed over raw experimental data that will establish a slight preference towards the morpheme structure phonotactics model.

4.4 Evaluation II: AUC score

The binary choice task of Becker and Gouskova (2016) motivates model evaluation based on the proportion of cross-class observation pairs they get correctly: if item x is accepted by one participant and item y is rejected by another participant, x should be ranked as more word-like than y.

Here, I present weighted AUC score computed by assigning each observation a weight $w_i = \frac{1}{N_i}$ where N_i is a number of observations for i-th nonce item (every nonce word contributes an equal mass of 1). The function $S(x, y)$ returns 1 if $f(x) > f(y)$, .5 if $f(x) = f(y)$, and 0 if $f(x) < f(y)$.

$$AUC = \frac{\sum_{x \in \mathcal{D}^+} \sum_{y \in \mathcal{D}^-} w_x w_y S(x,y)}{\sum_{x \in \mathcal{D}^+} w_x \sum_{y \in \mathcal{D}^-} w_y}$$

Table 2 suggests that the segment-based morphophonological model based on segmental-featural 2,3-cost phonotactic score outperforms alternatives. However, 95% CIs themselves do not allow to establish significant model differences. Similarly, the ROC AUC curve in Figure 4 does not show a clear superior model.

Model	AUC-ROC (item-weighted)
SF 2,3-cost (segment)	**.652** [.629, .673]
SF 2,3-cost (morpheme)	.640 [.617, .661]
MaxEnt (morpheme)	.637 [.615, .659]
Baseline	.613 [.592, .634]

Table 2: Item-normalized AUC score (larger is better, .5 is as good as chance). Brackets indicate 95% confidence intervals obtained by bootstrap resampling ($n = 5000$). Only models that outperform baseline are reported.

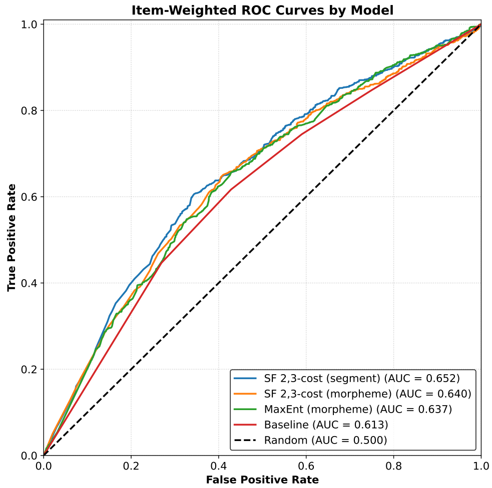

Figure 4: Item-weighted AUC-ROC curves by model.

To help distinguish between models, Table 3 presents bootstrapped differences in AUC score for the top model in Table 2 (namely, SF 2,3-cost; the implementation of the morpheme structure phonotactics model) and the rest of the models.

The bootstrapped differences suggest that the SF 2,3-cost model outperforms the alternatives: albeit the lower bound of the interval of difference between the MaxEnt model and the segmental SF 2,3-cost model is extremely close to zero.

Compared model	Mean Δ AUC [95% CI]
SF 2,3-cost (morpheme)	.012 [.006, .017]
MaxEnt (morpheme)	.014 [.002, .026]
Baseline	.039 [.021, .056]

Table 3: Pairwise differences in AUC scores between the segment-based segment-featural 2,3-cost model and other models. Values represent the mean difference (ΔAUC) with 95% confidence intervals in brackets, obtained via grouped bootstrap resampling ($n = 5000$). Intervals that do not include zero indicate a statistically significant difference at $\alpha = .05$.

Both evaluation metrics thus suggest that the segment-based morpho-phonological model, supplemented with an appropriate phonotactic score (the segmental-featural model of Durvasula 2026), predicts the productivity of Russian vowel-zero alternations at least just as well as the model presented by Becker and Gouskova (2016). Pairwise AUC differences provide a promising result in favor of the morpheme structure phonotactics model—but I leave the matter unsettled here.

4.5 Evaluation III: phonotactic tendencies

The morpheme structure phonotactics model (SF 2,3-cost) is able to account for the gradient productivity of Russian vowel-zero alternations. It's worth investigating how. In this section, I present observations regarding the behavior of the model with respect to the phonotactic tendencies discovered by Gouskova and Becker (2013).

4.5.1 Complex coda restriction

Nonce words that end in a CVCC syllable are less likely to undergo the vowel-zero alternation than nonce words that end in a CVC or CCVC syllable, reflecting the fact that while 17% of Russian masculine nouns end in CC# sequence, no alternating masculine nouns do.

For example, the nonce word *postor* is more likely to have the form *postr-a* than the nonce word *posotr* is to have the form *postr-a*. The model supports this: the former has one illicit UR n-gram (namely, /to$_{\text{F}}$r/) while the latter has two: /so$_{\text{F}}$t, o$_{\text{F}}$tr/.

Nonce word	Illicit segment 2,3-grams	SF 2,3-Cost
poso$_{F}$tr	/so$_{\text{F}}$t, o$_{\text{F}}$tr/	242
posto$_{F}$r	/to$_{\text{F}}$r/	1

Table 4: Illicit segment 2,3-grams and SF 2,3-cost for the nonce URs.

Note, however, that the complex coda UR *poso$_{F}$tr* violates 242 segment & feature n-grams while *posto$_{F}$r* only violates 1. The reason is due to the non-accidental complex coda ban in the lexicon. While 2 violated n-grams are segmental, all others are valued feature n-grams: the gap is natural class-based.

Nonce word	Illicit feature 2,3-grams	SF 2,3-Cost
poso$_{F}$tr	/$+flt$, $-syll$,$-syll$/ /$+flt$, $-son$,$+son$/ ...	242
posto$_{F}$r	—	1

Table 5: Illicit valued feature 2,3-grams and SF 2,3-cost for the nonce URs.

Consider, for example, that the nonce word *poso$_{F}$tr* includes an illicit valued feature n-grams /$+flt$, $-syll$,$-syll$/ and /$+flt$, $-son$,$+son$/: it suggests that there are no V$_{\text{F}}$CC sequences in the UR lexicon. As discussed in Section 2.3, this observation derives the complex coda restriction since all alternating vowels are the last vowels in their respective morphemes.

4.5.2 Sonority restrictions

Then, consider the sonority restrictions. nonce words that end in a TRVT syllable are less likely to undergo nonce words that end in a TTVR syllable. This reflects the fact that while 20% of Russian masculine nouns ends in a CR(V)C sequence, only .3% of alternating ones do.

For example, the nonce word *postor* is more likely to have the form *postr-a* than *posrot* is to have the form *posrt-a*. The nonce word *postor*, again, has as single illicit segmental n-gram, while *posrot* has two (although the absence of the /osr/ trigram is likely accidental and due to corpus).

Nonce word	Illicit segment 2,3-grams	SF 2,3-Cost
posro$_F$t	/osr, sro$_\mathrm{F}$/	16
posto$_F$r	/to$_\mathrm{F}$r/	1

Table 6: Illicit segment 2,3-grams and SF 2,3-cost for the nonce URs.

Importantly, the 14 illicit valued feature n-grams are all due to the /sro$_\mathrm{F}$/ segment trigram. They are not the /$+cons$, $+son$, $+flt$/ trigram: since the gap isn't complete.

Nonce word	Illicit feature 2,3-grams	SF 2,3-Cost
posro$_F$t	/$-voi$, $+son$,$+flt$/ /$+strid$, $+son$,$+flt$/ ...	16
posto$_F$r	—	1

Table 7: Illicit valued feature 2,3-grams and SF 2,3-cost for the nonce URs.

However, there are plenty of subcategories of consonants that do not occur in native CRV$_\mathrm{F}$ trigrams: for example, voiceless obstruents or stridents, providing a natural class-based reason for the model to disfavor nonce items that end in CRVC syllables as a group.

4.5.3 Monosyllabicity

The final restriction is monosyllabicity: monosyllabic items are less likely to undergo the alternation. This reflects the observation that while 8% of Russian masculine nouns are monosyllabic, only .7% of alternating ones are.

For example, the nonce word *pokor* is more likely to have the form *pokr-a* than the nonce word *kor* is to have the form *kr-a*. This is reflected in violated segmental n-grams: the nonce UR *poko$_F$r* has none, while the nonce UR *ko$_F$r* has one.

Nonce word	Illicit segment 2,3-grams	SF 2,3-Cost
ko$_F$r	/#ko$_\mathrm{F}$/	4
poko$_F$r	—	0

Table 8: Illicit segment 2,3-grams and SF 2,3-cost for the nonce URs.

And again, the segmental trigram /#ko$_\mathrm{F}$/ instantiates multiple natural class-based gaps: for example, all /#, $+dors$, $+flt$/ trigrams are absent from the UR lexicon—meaning that no monosyllabic alternating masculine nouns start with dorsals.

Nonce word	Illicit feature 2,3-grams	SF 2,3-Cost
ko$_F$r	/#, $+dors$, $+flt$/ /# $+high$,$+flt$/ ...	4
poko$_F$r	—	0

Table 9: Illicit valued feature 2,3-grams and SF 2,3-cost for the nonce URs.

Gaps in boundary initial #CV$_\mathrm{F}$ trigrams are an approximation of the restrictions on monosyllabic alternating nouns—since V$_\mathrm{F}$ must be the last vowel in the morpheme, words with a #CV$_\mathrm{F}$ trigram are monosyllabic alternating words. The relevant pattern in the data is that only 10 segmental #CV$_\mathrm{F}$ sequences are found in the UR lexicon (whereas there are 64 with non-floating mid vowels). Therefore, some monosyllabic nonce items will be penalized for #CV$_\mathrm{F}$ sequences and sequences of valued features they exemplify, leading to a lower acceptability of monosyllabic alternating nonce words.

4.5.4 Double-checking the effect of tendencies

To evaluate the effect of phonotactic tendencies, I ran Ordinary Least Squares (OLS) linear regression with the following structure. The model output for the nonce words served as the dependent variable, while the predictors included syllable count, complex coda, and sonority profile (whether the output of vowel deletion is TR, TTR, or RTR).

The predictors also included the number of consonants in the output of vowel deletion and the vowel itself, combined with an interaction factor between sonority sequencing and number of consonants (following the regression model reported by Becker and Gouskova 2016:404). To safeguard against potential heteroskedasticity and nonnormality in the model residuals, all parameter estimates were calculated using Heteroskedasticity-Consistent robust standard errors (HC3).

The results are reported in Table 10. The main conclusion is that all three phonotactic tendencies are significant predictors of the output of the proposed model (boldfaced).

Predictor	β	SE	z	p
Intercept	.167	.006	28.56	$<.001$
Monosyllabic	$-.018$	.003	-6.39	$<.001$
Complex coda	$-.030$	.002	-14.89	$<.001$
Vowel	$-.030$	.003	-8.87	$<.001$
Sonority profile	.037	.011	3.43	.001
Number of Cs	$-.132$	.005	-27.09	$<.001$
SP $\times$ # of Cs	$-.019$	.011	-1.74	.082

Table 10: OLS regression results probing the effect of phonotactic tendencies on the outputs of the morpheme structure phonotactics model.

The estimates show that the directionality of effects are correct—monosyllables and complex codas bring probability estimated by the proposed model down, whereas TTR and RTR outputs of vowel deletion raise it. I therefore conclude that the morpheme structure phonotactics model succeeds in capturing the effect of phonotactic tendencies on productivity of Russian vowel-zero alternations discovered by Gouskova and Becker (2013) and Becker and Gouskova (2016), along the lines discussed in Sections 4.5.1–4.5.3.

5 Conclusion & outlook

Nonce word tasks require careful consideration of the linking hypotheses between linguistic knowledge and performance in experimental settings. In this paper, I have pursued an explicit linking hypothesis between morpho-phonological nonce word tasks and "abstract phonology" approaches to morpho-phonological phenomena which rely on covert differences in underlying representations to achieve marking of phonological segments themselves as targets of exceptional phonological processes.

The core idea of the proposed morpheme structure phonotactics model is phonotactic evaluation of underlying representations (extending the ideas put forth by Ernestus and Baayen 2003 to abstract underlying segments). On top of that, the phonotactic evaluation itself is done using an n-gram counting model—suggesting that a weighted constraints model is not required for phonotactic wellformedness effects (Gorman 2013; Kostyszyn and Heinz 2021; Durvasula 2026).

The main result of this study is that the morpheme structure phonotactics model performs just as well as the sublexical MaxEnt model of Becker and Gouskova (2016)—and, perhaps, exceeds it—on their own nonce word judgement data regarding productivity of Russian vowel-zero alternations. The success isn't merely quantitative: as shown in Section 4.5, the phonotactic tendencies that characterize the masculine nouns that undergo the alternation are captured by a model which is only sensitive to the distribution of exceptional segments in the existing URs.

I believe the linking hypothesis to have two major conceptual predecessors in the literature: the traditional notion of morpheme structure constraints, negative statements about distribution of segments and their sequences in URs (Chomsky and Halle 1968; see the overview in Booij 2011 and see the recent defenses in Rasin and Katzir 2020; Gouskova 2023; Rasin 2025) and the strand conceptual criticism of phonotactic and morpho-phonological nonce word tasks, which takes the results to reflect static trends in the lexicon which are grammatically inert (Schütze 2005; Gorman 2013; Scheer 2019).

Whether this perspective extends to other alternations and languages remains a question for future work, but the present case study shows that morpho-phonological nonce word results are compatible with a wider range of phonological theories than previously thought. The findings thus neutralize the apparent conflict between morpho-phonological nonce word results and the empirical necessity of treating phonological segments themselves as targets or triggers of morpho-phonological processes (Martínez 2008; Rubach 2013, 2016; Round 2017; Zimmermann 2025 *i.a.*).

Acknowledgments

I thank Maria Gouskova, Kaustubh Ghoshal, Gillian Gallagher, Kyle Gorman, Adam Albright, and two anonymous reviewers for helpful feedback.

References

Adam Albright. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26(1):9–41.

Adam Albright and Bruce Hayes. 2003. Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition*, 90(2):119–161.

Blake Allen and Michael Becker. 2015. Learning alternations from surface forms with sublexical phonology. *Unpublished manuscript, University of British*

Columbia and Stony Brook University. Available as lingbuzz/002503.

Michael Becker and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in russian vowel deletion. *Linguistic inquiry*, 47(3):391–425.

Michael Becker, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language*, 87(1):84–125.

Geert Booij. 2011. Morpheme structure constraints. *The Blackwell companion to phonology*, pages 1–21.

Noam Chomsky and Morris Halle. 1968. *The sound pattern of English*. Harper&Row.

Huteng Dai, Connor Mayer, and Richard Futrell. 2023. Rethinking representations: A log-bilinear model of phonotactics. *Proceedings of the Society for Computation in Linguistics*, 6.

Karthik Durvasula. 2026. A closer look at what/how we can learn from computational modelling of phonotactics. *Ms. University of Michigan*.

Mirjam Ernestus and R Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79(1):5–38.

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

Maria Gouskova. 2012. Unexceptional segments. *Natural Language & Linguistic Theory*, 30(1):79–133.

Maria Gouskova. 2023. Mscs in positional neutralisation: the problem of gapped inventories. *Phonology*, 40(3-4):229–265.

Maria Gouskova and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory*, 31(3):735–765.

Maria Gouskova and Gillian Gallagher. 2020. Inducing nonlocal constraints from baseline phonotactics. *Natural Language & Linguistic Theory*, 38(1):77–116.

Morris Halle. 1959. *The Sound Pattern of Russian. A Linguistic and Acoustical Investigation*. 's-Gravenhage.

Morris Halle and Ora Matushansky. 2006. The morphophonology of russian adjectival inflection. *Linguistic inquiry*, 37(3):351–404.

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

Bruce Hayes, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. 2009. Natural and unnatural constraints in hungarian vowel harmony. *Language*, 85(4):822–863.

Michael Kenstowicz and Jerzy Rubach. 1987. The phonology of syllabic nuclei in slovak. *Language*, pages 463–497.

Kalina Kostyszyn and Jeffrey Heinz. 2021. Categorical account of gradient acceptability of word-initial polish onsets. In *Proceedings of the Annual Meetings on Phonology*.

Theodore McGraw Lightner. 1965. *Segmental phonology of modern standard Russian.* Ph.D. thesis, Massachusetts Institute of Technology.

Michal Temkin Martínez. 2008. Exceptionality and variation in modern hebrew spirantization. In *WCCFL 27*, pages 413–421.

Connor Mayer. 2025. Reconciling categorical and gradient models of phonotactics. *Society for Computation in Linguistics*, 8(1).

Connor Mayer and Max Nelson. 2020. Phonotactic learning with neural language models. *Society for Computation in Linguistics*, 3(1).

Scott Nelson. 2024. *The Computational Structure of Phonological and Phonetic Knowledge*. Ph.D. thesis, State University of New York at Stony Brook.

Joe Pater. 2008. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In *Phonological argumentation: Essays on evidence and motivation*. Equinox Publishing Ltd.

Ezer Rasin. 2025. Morpheme structure constraints solve three puzzles for theories of blocking in nonderived environments. *Linguistic Inquiry*, 56(4):713–749.

Ezer Rasin and Roni Katzir. 2020. A conditional learnability argument for constraints on underlying representations. *Journal of Linguistics*, 56(4):745–773.

Erich Round. 2017. Phonological exceptionality is localized to phonological elements: The argument from learnability and yidiny word-final deletion. *On looking into words (and beyond): Structures, relations, analyses*, pages 59–97.

Jerzy Rubach. 2013. Exceptional segments in polish. *Natural Language & Linguistic Theory*, 31(4):1139–1162.

Jerzy Rubach. 2016. Polish yers: Representation and analysis1. *Journal of linguistics*, 52(2):421–466.

Tobias Scheer. 2011. Slavic yers. In *The Blackwell companion to phonology*, pages 1–27. Wiley Online Library.

Tobias Scheer. 2019. On the difference between the lexicon and computation (regarding slavic yers). *Linguistic inquiry*, 50(1):197–218.

Carson T Schütze. 2005. Thinking about what we are asking speakers to do. *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 457–485.

Colin Wilson and Gillian Gallagher. 2018. Accidental gaps and surface-based phonotactic learning: A case study of south bolivian quechua. *Linguistic Inquiry*, 49(3):610–623.

Andrej Anatoljevich Zaliznjak. 1977. *Grammatičeskij slovar' russkogo jazyka [A grammatical dictionary of the Russian language]*. Moscow: Russkij Jazyk.

Eva Zimmermann. 2025. Cooperation at the morpho-phonology interface: An argument for phonological locality. In *Proceedings of the 41st West Coast Conference on Formal Linguistics*, pages 18–38.

Eva Zimmermann. 2026. Morphemes in phonology. In Paul de Lacy and Adam Jardine, editors, *The Cambridge Handbook of Phonology (2nd ed.)*. Cambridge University Press.

A Alternative phonotactic models

A.1 Accidental gaps

As mentioned earlier, Durvasula's (2026) model is not the only categorical phonotactic model on the market. There are arguably even simpler alternatives. For example, Gorman (2013) proposes a Boolean model that maps a nonce word to 0 if any of its n-grams are unattested in the corpus. Kostyszyn and Heinz (2021) propose a segmental cost model that maps a nonce word to how many of its n-grams are unattested in the corpus.

Both alternatives are subject to the *accidental gap* argument of Wilson and Gallagher (2018)—they are unable to distinguish between a genuine, natural-class based gap in the data, and an accidental gap. An illustration might help. Consider a corpus where voiced stops do not occur word-finally (*[$+voi$]#) and the segment /p/ does not occur non-finally.

Sequence	Boolean	S-Cost	S&F-Cost
at#	1	0	0
ap#	0	1	1
ad#	0	1	2
ab#	0	1	2

Table 11: Different scores in a toy example.

As discussed in Wilson and Gallagher (2018) for Quechua, a phonotactic model must distinguish between an accidental gap like **p#* and a real, natural class-based gap of **d#* and **b#*. Since both Boolean and segmental cost models only take attested segmental sequences into account, the difference is lost—as indicated by the scores. Only the segment-feature cost model is able to make the right cut: absence of the ⟨[$+voi$],#⟩ bigram adds onto the score of **d#* and **b#*. The discussion in Section 4.5.1. presents additional evidence in favor of sensitivity to natural class-based gaps.

The inherent issues of the alternative categorical models set in place suggest their inferior performance relative to both Becker&Gouskova's MaxEnt model and the morpheme structure phonotactics model supplemented with Durvasula's phonotactic scoring. As the next section shows, that is indeed the case.

A.2 Evaluation of alternative models

Here, I report the correlation metrics on all of the models: the numerical prefix refers to whether the model is sensitive to unattested bi- and tri-grams, while S cost refers to the segmental cost model of Kostyszyn and Heinz (2021) and Boolean refers to the Boolean model of Gorman (2013).

Model	Pearson r [95% CI]	Spearman ρ [95% CI]	Kendall τ_b [95% CI]
MaxEnt (morpheme)	**.893** [.837, .939]	.854 [.746, .915]	.673 [.566, .769]
SF 2,3-cost (segment)	.844 [.778, .925]	**.858** [.729, .928]	**.683** [.556, .796]
S 2,3-cost (segment)	.781 [.694, .874]	.792 [.630, .882]	.609 [.469, .724]
Baseline	.738 [.640, .825]	.655 [.480, .758]	.432 [.293, .560]
SF 2,3-cost (morpheme)	.713 [.605, .828]	.804 [.663, .885]	.607 [.478, .727]
S 2-cost (morpheme)	.668 [.511, .786]	.620 [.380, .762]	.425 [.243, .571]
S 2,3-cost (morpheme)	.656 [.534, .788]	.708 [.525, .815]	.500 [.348, .631]
2-Boolean (morpheme)	.656 [.495, .781]	.591 [.357, .741]	.402 [.218, .559]
3-Boolean (morpheme)	.656 [.493, .778]	.591 [.359, .742]	.402 [.219, .548]
SF 2-cost (morpheme)	.654 [.504, .777]	.609 [.390, .754]	.411 [.240, .561]
SF 2-cost (segment)	.552 [.344, .718]	.519 [.253, .716]	.373 [.184, .529]
S 2-cost (segment)	.357 [.059, .604]	.359 [.054, .606]	.250 [.041, .440]
2-Boolean (segment)	.273 [-.029, .546]	.286 [-.027, .565]	.214 [-.014, .431]
3-Boolean (segment)	.273 [-.029, .554]	.286 [-.037, .579]	.214 [-.023, .436]

Table 12: Syllabic type-level correlations between model predictions and empirical acceptability estimates. Brackets indicate 95% confidence intervals obtained by bootstrap resampling ($n = 5000$).

No alternative model outperforms the baseline (which is based on the mean reported wellformedness of the nominative bases in the study). For completeness' sake, I will also report the AUC scores of the S 2,3-cost (segment) model relative to the baseline, SF 2,3-cost, and MaxEnt model: the S 2,3-cost (segment) model is outpeformed by both SF 2,3-cost, and MaxEnt models. I conclude that natural class sensitivity is required for modelling the effects found in productivity of Russian vowel-zero alternations.

Model	AUC-ROC (item-weighted)
SF 2,3-cost (segment)	**.652** [.629, .673]
MaxEnt (morpheme)	.637 [.615, .659]
Baseline	.613 [.592, .634]
S 2,3-cost (segment)	.611 [.588, .632]

Table 13: Model comparison based on item-normalized AUC score (larger is better, .5 is as good as chance). Brackets indicated 95% confidence intervals obtained by bootstrap resampling ($n = 5000$).

B On feature decompositions

The feature decomposition employed in this paper is shown in Table 16 (next page). The core decision was to treat the $\pm float$ contrast as privative. This appendix describes the results of simulations with other feature decompositions, focusing on the AUC score of the SF 2,3-cost model (Binary-Flt). I entertain three additional possibilities: treating *float* as privative (only V_F is marked as $+$: Priv-Flt-Plus), treating *float* as privative (only regular segments are marked as $-$: Priv-Flt-Minus), and dispensing with *float*, marking V_F as underspecified for syllabicity (Priv-Syll).

Model	AUC-ROC (item-weighted)
Binary-Flt (SF 2,3-cost)	.652 [.629, .673]
Priv-Flt-Plus	.652 [.629, .673]
Priv-Flt-Minus	.643 [.620, .665]
Priv-Syll	.643 [.620, .664]

Table 14: Comparison of SF 2,3-cost models with different feature matrices based on item-normalized AUC score. Brackets indicated 95% confidence intervals obtained by bootstrap resampling ($n = 5000$).

The conclusion is that the privative / binary choice has no effect. What does have an effect is relative *markedness* of the vowel undergoing the alternation. If there is no feature that distinguishes it from the rest—model performance drops.

In that light, another worthwhile comparison is to see how big is the effect of the *segmental* part of the model—since the UR segment does provide another way to distinguish undergoing vowels from the rest. I report AUC scores where both UR and SR wellformedness is only evaluated relative to valued feature n-grams, for all four feature decompositions (thus, they are F 2,3-cost models).

Model	AUC-ROC (item-weighted)
Binary-Flt (F 2,3-cost)	.648 [.627, .670]
Priv-Flt-Plus	.648 [.627, .670]
Priv-Flt-Minus	.615 [.595, .635]
Priv-Syll	.615 [.596, .634]

Table 15: Comparison of F 2,3-cost models with different feature matrices based on item-normalized AUC score. Brackets indicated 95% confidence intervals obtained by bootstrap resampling ($n = 5000$).

Interestingly, omitting segment identity does not lead to a major drop in model performance. I leave further exploration of this finding for further work—I only note here that whatever the feature decomposition, it must be able to provide a natural class that only includes the vowels that undergo vowel deletion.

Segment	syll	cons	approx	son	cont	delrel	lateral	nasal	voice	dor	lab	cor	strid	ant	retro	high	low	back	stress	float	boundary
p	−	+	−	−	−	−	−	−	−	0	+	0	0	0	0	0	0	+	0	−	−
p^{j}	−	+	−	−	−	−	−	−	−	0	+	0	0	0	0	0	0	−	0	−	−
b	−	+	−	−	−	−	−	−	+	0	+	0	0	0	0	0	0	+	0	−	−
b^{j}	−	+	−	−	−	−	−	−	+	0	+	0	0	0	0	0	0	−	0	−	−
m	−	+	−	+	−	0	−	+	+	0	+	0	0	0	0	0	0	+	0	−	−
m^{j}	−	+	−	+	−	0	−	+	+	0	+	0	0	0	0	0	0	−	0	−	−
f	−	+	−	−	+	+	−	−	−	0	+	0	0	0	0	0	0	+	0	−	−
f^{j}	−	+	−	−	+	+	−	−	−	0	+	0	0	0	0	0	0	−	0	−	−
v	−	−	+	+	+	0	−	−	+	0	+	0	0	0	0	0	0	+	0	−	−
v^{j}	−	−	+	+	+	0	−	−	+	0	+	0	0	0	0	0	0	−	0	−	−
k	−	+	−	−	−	−	−	−	−	+	0	0	0	0	0	+	−	+	0	−	−
k^{j}	−	+	−	−	−	−	−	−	−	+	0	0	0	0	0	+	−	−	0	−	−
g	−	+	−	−	−	−	−	−	+	+	0	0	0	0	0	+	−	+	0	−	−
g^{j}	−	+	−	−	−	−	−	−	+	+	0	0	0	0	0	+	−	−	0	−	−
x	−	+	−	−	+	+	−	−	−	+	0	0	0	0	0	+	−	+	0	−	−
x^{j}	−	+	−	−	+	+	−	−	−	+	0	0	0	0	0	+	−	−	0	−	−
t	−	+	−	−	−	−	−	−	−	0	0	+	−	+	−	0	0	+	0	−	−
t^{j}	−	+	−	−	−	−	−	−	−	0	0	+	−	+	−	0	0	−	0	−	−
d	−	+	−	−	−	−	−	−	+	0	0	+	−	+	−	0	0	+	0	−	−
d^{j}	−	+	−	−	−	−	−	−	+	0	0	+	−	+	−	0	0	−	0	−	−
ts	−	+	−	−	−	+	−	−	−	0	0	+	+	+	−	0	0	+	0	−	−
s	−	+	−	−	+	+	−	−	−	0	0	+	+	+	−	0	0	+	0	−	−
s^{j}	−	+	−	−	+	+	−	−	−	0	0	+	+	+	−	0	0	−	0	−	−
z	−	+	−	−	+	+	−	−	+	0	0	+	+	+	−	0	0	+	0	−	−
z^{j}	−	+	−	−	+	+	−	−	+	0	0	+	+	+	−	0	0	−	0	−	−
ʂ	−	+	−	−	+	+	−	−	−	0	0	+	+	−	+	0	0	+	0	−	−
ʐ	−	+	−	−	+	+	−	−	+	0	0	+	+	−	+	0	0	+	0	−	−
tɕ	−	+	−	−	−	+	−	−	−	0	0	+	+	−	−	0	0	−	0	−	−
ɕ	−	+	−	−	+	+	−	−	−	0	0	+	+	−	−	0	0	−	0	−	−
ʑ	−	+	−	−	+	+	−	−	+	0	0	+	+	−	−	0	0	−	0	−	−
n	−	+	−	+	−	0	−	+	+	0	0	+	−	+	−	0	0	+	0	−	−
n^{j}	−	+	−	+	−	0	−	+	+	0	0	+	−	+	−	0	0	−	0	−	−
r	−	+	+	+	+	0	−	−	+	0	0	+	−	+	−	0	0	+	0	−	−
r^{j}	−	+	+	+	+	0	−	−	+	0	0	+	−	+	−	0	0	−	0	−	−
l	−	+	+	+	+	0	+	−	+	0	0	+	−	+	−	0	0	+	0	−	−
l^{j}	−	+	+	+	+	0	+	−	+	0	0	+	−	+	−	0	0	−	0	−	−
j	−	−	+	+	+	0	0	−	+	0	0	+	−	−	−	+	0	−	0	−	−
i	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	+	−	−	−	−	−
e	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	−	−	−	−
a	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	+	+	−	−	−
o	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	+	−	−	−
u	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	+	−	+	−	−	−
í	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	+	−	−	+	−	−
é	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	−	+	−	−
á	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	+	+	+	−	−
ó	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	+	+	−	−
ú	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	+	−	+	+	−	−
o$_{F}$	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	+	−	+	−
e$_{F}$	+	−	+	+	+	0	0	−	+	0	0	0	0	0	0	−	−	−	−	+	−
#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+

Table 16: Phonological feature matrix employed to calculate phonotactic scores in a Durvasula (2026)-like model.