Volume 29(2), August 31, 2023, 271-295
DOI: 10.17959/sppm.2023.29.2.271

The Phonology-Morphology Circle of Korea

Studies in Phonetics, Phonology and Morphology

ISSN: 1226-8690 / e-ISSN 2671-616X
http://www.phonology.or.kr

Modeling a phonotactic approach to segment recovery: The case of Japanese high vowels

James Whang (Seoul National University)*
Kakeru Yazawa (University of Tsukuba)

Abstract

Japanese listeners often have difficulties perceiving consonant clusters accurately and report hearing a vowel between the consonants, despite Japanese speakers producing numerous consonant clusters that result from a highly productive high vowel devoicing process. This poses a substantial challenge for phonological learning, as the task is to learn the strong CVCV preference in Japanese based on surface consonant clusters that violate this very preference. The current study investigates this learnability issue by building a computational model that induces Optimality Theoretic (OT) phonotactic constraints based strictly on overt speech. Two versions of the model are tested: one with classic OT-style faithfulness constraints that penalize violations, and one with positive constraints that reward sequences conforming to the constraints. Both have traditional markedness constraints. The results show that positive constraints are superior to faithfulness constraints in modeling phonotactic repair.

Keywords

acquisition, computational phonology, Japanese, perceptual repair

* E-mail: jamesw@snu.ac.kr
Received: June 22, 2023; Revised: July 28, 2023; Accepted: August 9, 2023

1. Introduction

The purpose of this paper is to provide a computational account of how prelexical infants might learn to "recover" a vowel that is not physically present in the acoustic input they receive (i.e., the caretaker's speech), focusing on Japanese high vowels. The model being proposed is called the statistical learning and repair model (STAR). Japanese high vowels present a unique learnability problem because while Japanese is argued to have a strong preference for CVCV structure (Shibatani 1990), a highly productive high vowel devoicing process that often results in the complete loss of vowel gestures (Shaw and Kawahara 2018) creates numerous consonant cluster-like sequences in actual speech that violate this phonotactic restriction (e.g. *suki* → *ski* 'to like'). Japanese infants therefore must learn that an apparent C_1C_2V sequence is equivalent to a $C_1\{i/u\}C_2V$ sequence by recovering the devoiced or deleted high vowel through repair (e.g., C_1 $C_2V \rightarrow C_1\{i/u\}C_2V$). Japanese infants as young as 12 months begin exhibiting insensitivity to C_1C_2V vs. C_1VC_2V distinctions (Kajikawa et al. 2006, Mugitani et al. 2007), suggesting that the repair process must be learnable pre-lexically at least in part. STAR is a first-pass attempt to model the acquisition of this repair process without a lexicon, and shows that positively defined constraints that reward rather than penalize certain sequences might be beneficial in phonotactic repair.

The architecture of STAR is largely based on the statistical learning and generalization model (STAGE; Adriaans and Kager 2010), which was originally proposed to stimulate unsupervised word segmentation in pre-lexical infants. STAGE by design does not have a lexicon that provides feedback for the model; as such, STAR also assumes no lexicon while learning to recover high vowels by essentially reversing the Japanese high vowel devoicing/deletion process and recover high vowels. In addition, STAR uses the same frequency-driven constraint induction (FDCI) mechanism of STAGE, but the induced phonotactic constraints are used to repair the loci of phonotactic violations with an appropriate vowel, rather than simply break up an illicit sequence.

Because STAR is a perception model, the model takes the overt form (symbolic representation of structureless, acoustic signal) as input and gives what the model thinks is the corresponding surface form (what Japanese listeners think they heard) as output. Tesar and Smolensky (1996) decompose the learning problem by first proposing three levels of representation that are relevant in phonological learning: [overt], /surface/, and |underlying| forms. The mapping from [overt] to /surface/ form corresponds to perception, while the mapping from /surface/ to |underlying| form corresponds to word recognition (Apoussidou 2007). The overt form of a word is the concrete acoustic signal produced by a speaker. As it is purely an acoustic signal, it carries no inherent linguistic structure. Structure is added in the surface form by the listener based on one's grammar. At this stage, the listener is yet to recognize the form as a word. To recognize a word, the listener must map the surface form to a matching underlying form in the lexicon. To give a more concrete example, a Japanese infant must learn to take an overt form like [s_ki] that begins with a cluster-like sequence, recover the devoiced high vowel as in /suki/ at the surface level, then arrive at the underlying representation |suki| 'to like'. Since STAR is lexiconless, there is no underlying form to speak of, and thus the model takes (phonetically transcribed) [overt] forms as input and returns /surface/ forms as output. In other words, it models what a Japanese phonotactic grammar would "hear" given an acoustic input.

The paper begins with a literature review in section 2 on how the issue of constraint-based phonotactic acquisition has been treated within the Optimality Theoretic framework. The theoretical motivations for positively defined constraints as well as the experimental literature on Japanese high vowel production and perception are also discussed. Section 3 describes STAR in detail, starting with the data used to train the model and the process of how the model selects what it considers the most harmonic output. Section 4 gives simulation results that compare two versions of STAR, one with positive constraints and another with traditional faithfulness constraints that penalize violations. Section 5 concludes the paper with a summary and suggestions for future research.

2. Background

2.1 Constraint-based phonotactic acquisition

In a constraint-based phonological framework, language learning essentially boils down to learning the correct constraint ranking. Most constraint-based frameworks assume innate constraints (e.g., Constraint Demotion Mechanism, Tesar 1995; Gradual Learning Algorithm, Boersma and Hayes 2001), but phonotactic learning models such as the maximum entropy model (MaxEnt; Hayes and Wilson 2008) successfully show that the constraints can be induced. MaxEnt assumes that Universal Grammar is composed of an innate set of features and a format for markedness constraints. Given a feature set, the model induces all logically possible phonotactic constraints. From this space of possible constraints, the model selects and ranks the constraints in a way that maximizes the probability of the input data with a built-in preference for general constraints over specific constraints. The resulting constraint set therefore is language-specific since constraints that were induced but are inactive in the given language are discarded. Because the MaxEnt model induces constraints based on a given feature set, the constraints in effect exist independently of any input data it receives. The model, therefore, explicitly does away with a universal constraint set (CON), which is a notable divergence from previous approaches. However, since the feature set is part of Univeral Grammar (UG), the result is still a universal constraint space which language learners must tap into to extract a language-specific constraint set.

Adriaans and Kager (2010) propose the statistical learning and generalization model (STAGE), which also induces constraints based solely on the input. While STAGE assumes a universal feature set, unlike MaxEnt it has no built-in format for constraints and thus steps away from the notion of CON even further. Instead, STAGE employs frequency-driven learning and generalization mechanisms to induce constraints that are relevant for the data. Each constraint is assigned a weight based on the statistical distribution of the biphones the contraint is concerned with. As will be discussed in more detail in section 3, STAGE induces markedness constraints for underrepresented (dispreferred) biphones and faithfulness constraints for overrepresented (preferred) biphones. From the biphone-specific constraints, more general constraints are formed via single-feature abstraction. For example, suppose the model induces the constraints CONT-(bl) and CONT-(pl). These constraints are different by a single feature, namely *voice*. The generalization mechanism therefore takes these constraints and forms a more general constraint CONT-(x∈{p,b};y∈{l}), which says, "Assign a violation for every *pl* or *bl* sequence in the input that is not preserved in the output." The model additionally showed that with the statistical and generalization mechanisms, a lexicon is not required for successful phonotactic learning but rather that phonotactic knowledge can help lexical acquisition.

STAR takes the approach of STAGE (Adriaans and Kager 2010) and does not assume an *a priori* representation of phonotactic constraints. Since STAGE showed that a lexicon is not necessary for phonotactic learning to take place, STAGE follows suit and also has no lexicon.

2.2 Positive constraints

STAR also explores the possible role of positively defined constraints that reward rather than penalize certain sequences (hereafter, positive constraints) in a phonotactic grammar. The theoretical motivation behind the induction of positive constraints comes from a usage-based approach to phonology (Bybee 2006). Distinct elements that frequently pattern together have been shown to be treated functionally as a single unit by the grammar, and positively defined constraints that can link and reward certain overrepresented sequences (in this case phones) provide the mechanism necessary to do so (Välimaa-Blum 2009). For example, /ç/ has an extremely high co-occurrence rate with /i/ in Japanese, and a positive constraint like +çi that says, "Assign a reward for every instance of *çi* in the output," grammaticalizes this pattern straightforwardly. This means that when the model encounters an input like [çto] 'person', it can epenthesize /i/ in the output to form /çito/ to satisfy the constraint. The applicability of positive constraints is not limited just to the usage-based approach. They have also been shown to be advantageous in driving autosegmental spreading (e.g., (+)Spread-F(eature): Assign a reward for each segment linked to F as a dependent; Kimper 2016).

Because the approach taken here is dependent on language use, constraints are assumed to be induced based on positive occurrences in the language rather than possible but never occurring phone combinations (Välimaa-Blum 2009). This characterization of learning strictly based on positive evidence seemingly stands in opposition to what Tesar and Smolensky (1998) call the utilization of *implicit negative evidence*. Tesar and Smolensky (1998) argue that given the OT framework, every instance of positive evidence necessarily results in the rejection of all possible competitors generated by GEN. The infant is assumed to have access to the rejected candidates, which essentially form implicit but positive instances of negative evidence. Assuming innate constraints fully equips the learner to pick up on systematic, consistent gaps in the input data, from which the learner can conclude that these forms must be prohibited in their language (Hayes 1999).

The question that seems seldom asked, however, is at what point the learner can arrive at the conclusion that unencountered forms are indeed prohibited. To state differently, how much implicit negative evidence is enough? This issue is especially relevant to models like STAR that take a more input-centric approach to phonotactic learning and do away with innate constraints. Since constraints are induced based solely on encountered forms, the learner has no way of dealing with unencountered forms. As the simulations in section 4 will show, for an input-driven learning algorithm, a mechanism that can grammaticalize frequently co-occurring units (e.g., positive constraints) is useful for learning sequences that are preferred in the language in the absence of sufficient implicit negative evidence.

2.3 Experimental work on Japanese high vowels

In a well-known series of studies by Dupoux et al. (1999, 2011), French and Japanese speakers were presented with acoustic stimuli with the high back rounded vowel [u] of varying durations ranging from 0 ms to 90 ms occurring between two consonants (e.g., [ebzo] → [ebuːzo]). The stimuli were designed so that when there is no vowel in the stimuli, the result is a sequence that is phonotactically legal in French but illegal in Japanese. Their results showed Japanese speakers were unable to distinguish vowel-ful vs. vowel-less tokens, erring heavily towards perceiving both as vowel-ful. The authors propose that the results are due to the prohibition against heterorganic consonant sequences in Japanese phonotactics and that a 'top-down' phonotactic effect on perception forces such illegal sequences to be repaired automatically as the nearest legal sequence, namely one that epenthesizes a minimal vowel to break up such sequences.

Contrary to what one might expect from the perception patterns above, Japanese speech actually contains many heterorganic consonant sequences. During production, speakers of standard modern Japanese at the least devoice the high vowels /i/ and /u/ when the vowels are unaccented and flanked by two voiceless obstruents. For example, while the /u/ in /kúɕi/[1] 'free use' and /kuɕi/ 'skewer' are both between two voiceless consonants, only /kuɕi/ 'skewer' undergoes devoicing because the vowel is unaccented. Likewise, the /u/ is unaccented in both /kuki/ 'stem' and /kugi/ 'nail', but only /kuki/ 'stem' undergoes devoicing because the /u/ is flanked by two voiceless stops, namely /k/. Although the process is more commonly called high vowel devoicing in the literature, the process actually results in a range of articulatory effort, from simple loss of voicing to complete deletion of the vowel (Shaw and Kawahara 2018), and thus the term 'reduction' is used hereafter to encompass both devoicing and deletion. Crucially for the current study, Whang (2018) showed that when /i, u/ reduce, whether the reduction results in simple devoicing or complete deletion is not random, but rather

[1] Note that underlying |s| neutralizes to /ɕ/ before /i/ in Japanese.

conditioned on the predictability of the reducing vowel in a given context. Highly predictable vowels tend to delete while less predictable vowels tended to devoice, leaving behind coarticulatory traces in the burst/frication of the preceding consonant.

In the experimental evidence is a puzzle. Japanese speakers consistently produce numerous, heterorganic, voiceless consonant sequences as a consequence of high vowel reduction. However, Japanese listeners show a dispreference for such sequences and repair them by epenthesizing a high vowel. One might argue that the perceptual repair is simply recovering a vowel that is lost due to a productive and predictable process. However, Ogasawara and Warner (2009) found in a lexical judgment task that when Japanese listeners were presented with unreduced forms of words where reduction is typically expected, reaction times were longer than when presented with reduced forms. This suggests that reduced forms, despite their phonotactic violations, can have a facilitatory effect on lexical access due to their commonness, making vowel recovery actually unnecessary (Cutler et al. 2009, Ogasawara 2013). In other words, CC sequences are not perceptually equivalent to CVC sequences to Japanese listeners. Furthermore, Japanese infants also begin to show a noticeable decline in discriminating CC vs. CVC sequences by the age of 1 year; 0 months (Kajikawa et al. 2006, Mugitani et al. 2007), at a stage when they should know only a few words (Kalashnikova et al. 2016), making it difficult to argue that the CC repair to CVC is a lexicon-driven effect. The current paper, therefore, models the perceptual repair process of Japanese listeners from a purely phonotactic perspective, exploring how Japanese listeners come to exhibit a CVCV preference despite learning the phonotactic structure of the language from input that contains numerous violations.

3. The Statistical Learning and Repair Model (STAR)

3.1 Input for the model

As is generally the case with phonotactic models, STAR is trained and tested on phonetically transcribed data. The input for the model assumes a segmental rather than a featural representation, allowing the model to explore how far learners can get with phonemes alone. A "stage of vulnerability" is also assumed, where an infant has yet to realize that certain phones are in allophonic relationships and thus treat them all as separate phonemes (Hayes 2004). The symbolic representation for the input to STAR is based on the results from Whang (2018), which suggest that when high vowels reduce, the variability between deletion and reduction is not arbitrary but conditioned by how predictable the reducing vowel is in a given consonant. The predictability of a vowel is associated with whether only one or both of /i, u/ can occur after a given consonant, which is summarized in Table 1 below.

Table 1: Consonants and possible following vowels according to high vowel predictability. △ means vowel allowed in limited cases (e.g., loans).

		i	u	non-high
Unpredictable	p	○	○	○
	k	○	○	○
	ɕ	○	○	○
Predictable	ʦ	×	○	×
	ʨ	○	△	○
	ɸ	△	○	△
	ç	○	△	○
	s	×	○	○

Whang (2018) excluded /p/ due to its overall rarity (Tsuji et al. 2014) and /ʦ/, which can only be followed by /u/, unlike other consonants that generally allow non-high vowels to follow as well. Acoustic analyses showed that predictable high vowels tended to delete while they devoiced in non-predictable contexts, leaving behind traces of the coarticulated vowel.

Since in the unpredictable cases (i.e., [k, ɕ] and by extension [p]) the consonant burst/frication noise includes coarticulatory information about the following vowel, this means that words like /kiɕi/ ‘a coast’ and /kuɕi/ ‘a skewer’ can reliably distinguished even if the first vowel is reduced because the coarticulatory information is retained in the /k/ (i.e., [k^{i}ɕi] versus [k^{u}ɕi]).[2] This coarticulatory information is also present when the following vowel is not reduced (e.g., /kizu/ → [k^{i}izu] ‘a wound’; also /ɕi/ → [$ɕ^{i}$i] in the two previous examples). In terms of the model, this means that words like the three examples given above can be symbolically represented as the

[2] Note that consonants with phonetic /i, u/ coarticulation are assumed to be distinct from consonants with phonemic secondary articulations (e.g., [k^{i}] ≠ [k^{j}]) following the conventions of Japanese phonological literature. Whether these segments actually differ acoustically remains to be investigated.

following: <k^{i} ɕ^{i} i> <k^{u} ɕ^{i} i> <k^{i} i z u>. For the predictable cases (i.e., [ɸ, s, ç, t͡ɕ] and by extension [t͡s]), the consonants were shown to not carry the vowel information, so a word like /çito/ would be produced as [çto], which in turn means that the input for the model would also look like <ç t o>.

One million pseudo-Japanese words that were two to five morae in length were generated to serve as the model's training data. These 'words' were generated with a word generator that we built based on Japanese phoneme transition probabilities, which were calculated from a 288 million word lexical corpus of Asahi Newspaper printed between 1985 and 1998 (Tamaoka and Makioka 2004). The benefits of using a generator lie in its flexibility and ability to generate as many words as necessary that are realistic (and often indeed real) without having to rely on linguistic intuition. Examples of the generated pseudo words are provided in Appendix A.

There are three issues with the corpus data, however, that warrant some discussion. First, because the corpus was phonemically transcribed based on orthographic forms, the pseudo-words generated were converted to reflect overt forms. This conversion was necessary because the input for an infant learner is the output of the infant's caretaker, which presumably has undergone normal phonological processes of the language, including high vowel reduction. As briefly mentioned above, coarticulatory information was symbolically represented in the data for cases in which devoicing rather than deletion is likely to occur (i.e., unpredictable cases). For example, [ɕi] was represented as <ɕ^{i} i> (i.e., i-coarticulated ɕ followed by i). This essentially makes all high vowels predictable after a given C_1. While this may seem redundant at first glance, it has been shown that adult Japanese speakers can reliably predict which high vowel will occur based solely on the acoustic information available in i- and u-coarticulated [ɕ] (Beckman and Shoji 1984). In addition, this coarticulation effect is reported to be present even when no full vowel follows the consonant (Whang 2018). If it is the case that adults are sensitive to the acoustic difference of these sounds, it does not seem unreasonable to assume that infants may be sensitive to them as well.

Second, the transitional probabilities reported in Tamaoka and Makioka (2004) did not take into account word boundaries. The study looked simply at how frequent particular phonemes were in the corpus. Bimoraic transitional probabilities were also reported. Because word boundary information was absent in the study, the words generated by the fake word generator also did not include word boundaries. This in turn means that the models did not induce any constraints that involve word boundaries either.

Third, since only unaccented high vowels get reduced and pseudo-words do not have lexical pitch accents, the rate in which reduction applies when a high vowel is surrounded by two voiceless obstruents had to be estimated. An estimate of 55% was taken from Maekawa and Kikuchi (2005), which was a large corpus-based study of high vowel reduction rates in spontaneous speech. The same study showed that vowels reduced the least when flanked by fricatives, and thus the reduction rate presented in the Maekawa and Kikuchi paper were also applied to the generated data.

3.2 Learning the grammar

The statistical portion of STAGE, which STAR is based on, calculates the observed/expected ratios (O/E) of all biphones that occur in the input data to induce phone-specific constraints when an assumed threshold is reached. O/E ratios compare how often a biphone actually occurs in the data (observed) to how often that biphone should have occurred if all segments had equal likelihoods of combining (expected). The O/E ratio for a given biphone *xy* is calculated by dividing the probability of *xy* by the product of the summed probability of all biphones beginning with *x* and the summed probability of all biphones ending with *y,* as shown in the equation below. The resulting value quantifies the magnitude of the biphone's over-/under-representation in the data. An O/E of 1.0 indicates that a biphone occurred exactly as often as expected. An O/E of 0.5 indicates that a biphone occurred half as often as expected, and an O/E of 2.0 indicates that a biphone occurred twice as often as expected.

$$\frac{O(xy)}{E(xy)} = \frac{Pr(xy)}{\sum Pr(xY) * \sum Pr(Xy)}$$

The frequency-driven constraint induction mechanism (FDCI) of STAGE induces a markedness constraint against a biphone when the O/E of the sequence is less than 0.5 (**xy*); and when the O/E is higher than 2.0, FDCI induces a contiguity constraint that keeps a matching contiguous sequence in the input also contiguous in the output (e.g., CONT-*yz*. The markedness and contiguity constraints are defined as follows:

- **xy*: Assign a violation for every *xy* sequence in the output.

- CONT-*yz*: Assign a violation for every *yz* sequence in the input that is not preserved in the output.

STAR induces and ranks constraints in the same way as STAGE, but there are two main differences between the two models. First, whereas STAGE induces contiguity constraints for overrepresented biphones, the very same constraints in STAR are redefined positively as constraints that operate strictly on the output, rewarding matches rather than penalizing violations, defined as follows:

- +*yz*: Assign a reward for every *yz* sequence in the output.

Note that the choice to look at biphones rather than a longer sequence was not for the sake of simplicity but because in perception, the information necessary to recover a reduced vowel is wholly contained in the consonant preceding it. By focusing on biphones, the model can learn that a certain consonant C_1 frequently cooccurs with a certain vowel V_1, and also that C_1C_2 clusters are rare. Therefore, when the model encounters a word that contains a C_1C_2 cluster, it only needs C_1 to decide that V_1 should be inserted. The role that C_2 plays is simply to inform the model of the location in which the target vowel is to be inserted. In other words, while it is true that high vowel reduction is conditioned by the voicelessness of the two consonants flanking the target vowel for production, the consonant following the vowel is irrelevant for the purposes of perception.

The second difference is that in STAR, the constraints are assumed to be weighted, as in serial Harmonic Grammar (sHG; Pater 2012), rather than ranked in strict domination, as in Optimality Theory (OT; Prince and Smolensky 1993/2004). Therefore, the expected values of the biphones are used as weights in STAR rather than a means to ordinally rank the constraints as in STAGE. The sHG framework is adopted to restrict GEN to performing one operation at a time to avoid the *infinite goodness* problem that is inherent to positive constraints (Prince 2007, Kimper 2016). Weighted constraints, therefore, were used for the sake of consistency with the general sHG framework and not based on a commitment to a particular interpretation of the constraints' numerical values.

3.3 Modeling perceptual repair

Having learned the grammar based on the training data, the model is presented with test data and given the task of deciding whether phonotactic repair is required, and if so how. Based on a given input, a GEN mechanism generates a faithful candidate and all possible forms which may or may not be grammatical. The only restriction is that the input is always assumed to be the correct overt form (i.e., perfect noiseless perception), and thus every segment retains its identity. The model therefore calculates the most harmonic surface form that would be pronounced as the overt form (Smolensky 1996, *et seq.*). Also, GEN can only generate candidates that differ from the input by only one segment (*à la* sHG), and thus the candidate set is finite. GEN performs one of two operations: delete or epenthesize a segment (consonant or vowel). Next, an EVAL mechanism takes the candidates generated by GEN and returns the most optimal surface form. Since the constraints are weighted, the form that has the highest sum of constraint values is chosen as the winner. The optimal form is then fed back into GEN for another set of candidates with one-segment differences, which is then evaluated again by EVAL. This GEN → EVAL loop is repeated until the derivation converges on a form that can no longer be improved with single segment operations.

4. Simulation

STAR was tested on its ability to epenthesize the correct high vowel when it encountered a consonant cluster. How the models dealt with unreduced forms was also tested. If the model learned the correct constraint set and ranking, it should return the same output for the reduced and unreduced forms (e.g., [ski, su̥ki, suki] → /suki/ 'like'). As discussed in section 2, STAR uses positive constraints to deal with sequences that were not encountered during training. STAR is compared directly to a STAGE-like model that induces contiguity constraints. The model being compared is not STAGE *per se*, as STAR lacks a constraint generalization mechanism and uses FDCI only. To avoid confusion, the STAGE-like version of the model hereafter is called the *continguity* statistical learning and repair model (C-STAR) and the model being proposed here, the *positive* statistical learning and repair model (P-STAR), since it induces positive constraints. Because the contiguity and positive constraints were induced by the same biphone sequences from the same training data, their weights (i.e., the biphones' expected values) were equal in the two models. The markedness constraints in the EVAL mechanism for both models penalize candidates for each violation.

Constraints that apply to VC sequences were excluded in the simulations to introduce a processing bias towards prioritizing CV sequences. Phonologists have long been aware of a crosslinguistic preference for CV sequences, and in Optimality Theory this preference was formalized in the form of an ONSET constraint that penalizes a syllable that lacks an onset and a NOCODA constraint that penalizes a syllable that has a coda. Together, the result is a preference for a CV structure. In addition, Endress and Bonatti (2007) argue that when processing speech on-line there are two mechanisms at play. The first mechanism rapidly extracts structural information (syllables) from speech, which is then fed to a slower mechanism that detects statistical regularities within the extracted structures. In other words, there is an innate linguistic restriction on what sequences statistical computations can be applied to. If this is indeed the case, it does not seem unreasonable to think that calculating distributional probabilities of CV is prioritized over VC sequences in linguistic systems.[3]

4.1 Materials

Examples of the training words can be found in Appendix A. To train the model, one million tokens of pseudo-Japanese words were generated by the word generator as described in section 3.1. The training words were two to five morae in length. Test data were 33 reduced-unreduced pairs of real words. For each of the 11 out of 12 possible voiceless obstruent-high vowel combinations, there were three words that began with the combination. Two of the three test words contained one target vowel (e.g., /ɸu̲ku/ 'clothes'; /ɸu̲ɕi/ 'knuckle'), whereas the third had two target vowels (e.g., /ɸu̲ɕi̲kateː/ 'motherless family'). This means that of the 33 pairs in the test data, 22 of them had one target vowel and the remaining 11 had two target vowels, for a total of 44 reducing environments. A complete list of the test words used can be found in Appendix B.

[3] Although not presented here in consideration of space, a separate simulation showed that a CV bias can be induced from the data anyway. This suggests that at least in the case of Japanese, assuming an innate CV bias may not be necessary.

4.2 Procedures

The FDCI mechanism induced constraints based on the training data. Both C-STAR and P-STAR induced a markedness constraint for biphone sequences with an O/E ratio ≤ 0.5. For the STAGE-like C-STAR, contiguity constraints were induced for biphone sequences with an O/E ratio ≥ 2.0. For P-STAR, the same O/E threshold of 2.0 was used to induce positive constraints. The two models were then given the test data and the task to determine whether vowel epenthesis is required for any given word, and if so what vowel. The models were evaluated on how they dealt with reduced and unreduced forms separately. The reduced forms were evaluated on whether a vowel was inserted between the target consonant clusters. If a vowel was inserted, it was then evaluated for the rate in which the appropriate vowel was inserted. For the unreduced forms, which should be favored by the CVCV bias of Japanese, the models were evaluated on whether any vowels were deleted or inserted despite the words needing none.

4.3 Results

The results for unreduced forms are reported first. Both C-STAR and P-STAR had a success rate of 100%. A survey of the constraints revealed that this is because for all CV biphones, the highest ranked constraints were for the overrepresented sequences (i.e., contiguity or positive constraints). This means that in C-STAR, the faithfulness constraints are responsible for keeping CV sequences in the input intact in the output. Likewise, in P-STAR, while the positive constraints do not care about the input, it still prefers the output candidates that contain the particular CV sequences.

The results for the reduced forms in the test data, however, show an advantage of P-STAR over C-STAR. First, when it comes to identifying the location in which vowel epenthesis is required, P-STAR is successful 100% of the time (44 of 44) , but C-STAR fails to epenthesize anything in three instances (i.e., in [ɕ^{u}k^{u}seː] 'purge', [ɸ k^{u}u] 'clothes', and [çto]'person'). Second, among the cases in which the two models correctly identified the contexts where epenthesis is expected, P-STAR outperforms C-STAR by a wide margin with a success rate of 89% (39 of 44) versus 17% (7 of 41). A breakdown of what vowels C-STAR and P-STAR inserted for each consonant is shown below in Tables 2 and 3, respectively.

Table 2: Vowels inserted for C_1 by C-STAR.

	Unpredictable						Predictable					
	/i/ target			/u/ target			/i/ target		/u/ target			
	p^{i}	k^{i}	ɕ^{i}	p^{u}	k^{u}	ɕ^{u}	tɕ	ç	ʦ	ɸ	s	
i	**-**	**1**	**1**	-	1	1	**1**	**0**	1	-	1	
u	2	1	-	**1**	**0**	**0**	2	-	**1**	**1**	**1**	
a	1	-	-	-	2	-	-	-	1	-	2	
e	-	2	2	-	-	1	2	1	-	1	-	
o	-	-	1	2	3	-	3	1	2	-	1	
Correct	0/3	1/4	1/4	1/3	0/6	0/3	1/5	0/3	1/5	1/3	1/5	**17%**

Table 3: Vowels inserted for C_1 by P-STAR. 100% accuracy except /s/.

	Unpredictable						Predictable					
	/i/ target			/u/ target			/i/ target		/u/ target			
	p^{i}	k^{i}	ɕ^{i}	p^{u}	k^{u}	ɕ^{u}	tɕ	ç	ʦ	ɸ	s	
i	**3**	**4**	**4**	-	-	-	**5**	**3**	-	-	-	
u	-	-	-	**3**	**6**	**3**	-	-	**5**	**3**	**-**	
a	-	-	-	-	-	-	-	-	-	-	5	
e	-	-	-	-	-	-	-	-	-	-	-	
o	-	-	-	-	-	-	-	-	-	-	-	
correct	3/3	4/4	4/4	3/3	6/6	3/3	5/5	3/3	5/5	3/3	0/5	**89%**

P-STAR consistently epenthesized /a/ after /s/ due to a high ranking +sa constraint being induced that was weighted higher than +su. This was due to /a/ being more frequent than any other vowel after /s/ in the Japanese training data.

For C-STAR, the failure to insert a vowel were due to two different reasons: (i) absence of a markedness constraint against particular consonant clusters and (ii) presence of CONT constraints that favor keeping certain consonant clusters intact. When the model did insert a vowel in the correct contexts, the choice of vowel was often wrong because the combination of CONT and markedness constraints could not arrive at the correct winner. First, the one case in which the lack of a markedness constraint was problematic is shown (i.e., [ɕ^{u} k^{u}seː] 'purge'). Table 4 below shows in detail how the model evaluated the word. In the first iteration for [ɕ^{u}k^{u}seː], the shortcomings of just having CONT and markedness constraints become clear. Here the high ranking constraint *k^{u}s successfully eliminates the faithful candidate /ɕ^{u}k^{u}se:/. However, the CONT constraints fail to pick out the winner from among the epenthesized candidates. The CONT constraints do nothing here because the input lacked a vowel to begin with, resulting in multiple candidates with the same total weight. The model is required to return an output regardless, so it selects a winner at random (the candidate that GEN happened to list first among the multiple candidates), which happens to be the wrong candidate /ɕ^{u}k^{u}ose:/. The winner of the first iteration is then used as the input in the second iteration. Here, none of the candidates are eliminated because the model induced no markedness constraint against /ɕ^{u}k^{u}/, making the faithful candidate equally good as any of the epenthesized candidates. Therefore, the model again picks a winner at random (i.e., /ɕ^{u}k^{u}ose:/).

Table 4: Evaluation of [ɕ^{u}k^{u}se:] 'purge' by C-STAR. Correct winner marked by ✔. Actual winner picked by model marked by ✗.

1 [ɕ^{u} k^{u} se:]	CONT-k^{u}u 0.001318	*k^{u}s 0.000510	CONT-se: 0.000247	CONT-ɕ^{u}u 0.000197	*Total Weight*
/ɕ^{u} k^{u} se:/		-1			-0.00510
/ɕ^{u} k^{u} ase:/					0
/ɕ^{u} k^{u} ise:/					0
✔ /ɕ^{u} k^{u} use:/					0
/ɕ^{u} k^{u} ese:/					0
✗ /ɕ^{u} k^{u} ose:/					0

2 [ɕ^{u} k^{u} ose:]	CONT-k^{u}u 0.001318	*k^{u}s 0.000510	CONT-se: 0.000247	CONT-ɕ^{u}u 0.000197	*Total Weight*
✗ /ɕ^{u} k^{u} ose:/					0
/ɕ^{u} ak^{u} ose:/					0
/ɕ^{u} ik^{u} ose:/					0
/ɕ^{u} uk^{u} ose:/					0
/ɕ^{u} ek^{u} ose:/					0
/ɕ^{u} ok^{u} ose:/					0

In contrast, shown below in Table 5 is how P-STAR evaluated the same word. The lack of the markedness constraint *ɕ^{u}k^{u} is not a problem for P-STAR because positive constraints are only concerned with the output. In the first iteration, the candidate

/ɕ^{u}k^{u}use:/ wins because in addition to satisfying the +se: constraint like all of the other candidates, it also satisfies the highest ranked constraint +k^{u}u. This candidate is then used as the input for the next iteration. Here, all candidates satisfy +k^{u}u and +se:, but /ɕ^{u}uk^{u}use:/ is chosen as the winner because it also satisfies the lowest ranked but nevertheless decisive +ɕ^{u}u constraint.

Table 5: Evaluation of [ɕ^{u}k^{u}se:] 'purge' by P-STAR. Clear winner.

1 [ɕ^{u} k^{u} se:]	+k^{u}u 0.001318	*k^{u}s 0.000510	+se: 0.000247	+ɕ^{u}u 0.000197	*Total Weight*
/ɕ^{u} k^{u} se:/		-1	+1		0.000263
/ ɕ^{u} k^{u} ase:/			+1		0.000247
/ ɕ^{u} k^{u} ise:/			+1		0.000247
☞ / ɕ^{u} k^{u} use:/	+1		+1		0.001565
/ ɕ^{u} k^{u} ese:/			+1		0.000247
/ ɕ^{u} k^{u} ose:/			+1		0.000247

2 [ɕ^{u} k^{u} use:]	+k^{u}u 0.001318	*k^{u}s 0.000510	+se: 0.000247	+ɕ^{u}u 0.000197	*Total Weight*
/ɕ^{u} k^{u} use:/	+1		+1		0.001565
/ɕ^{u} ak^{u} use:/	+1		+1		0.001565
/ɕ^{u} ik^{u} use:/	+1		+1		0.001565
☞ /ɕ^{u} uk^{u} use:/	+1		+1	+1	0.001762
/ɕ^{u} ek^{u} use:/	+1		+1		0.001565
/ɕ^{u} ok^{u} use:/	+1		+1		0.001565

Second, in the case of the words [ɸ k^{u}u] 'clothes' and [ç to] 'person', C-STAR failed to insert a vowel because there were eight CC sequences that were overrepresented in the data (<p^{i} ʦ> <ɕ^{i} k^{j}> <ɕ^{u} p^{j}> <ɕ^{u} ʦ> <ɸ k^{u}> <ç t> <ç ʦ> <ç k^{i}>), which resulted in the induction of contiguity constraints. Since the two words each involved one of these sequences, namely <ɸ k^{u}> and <ç t>, the faithful candidate was incorrectly chosen as the winner. Table 6 below is an example of how the word [çto] 'person' was evaluated by C-STAR. Since <ç t> happened to be one of eight CC sequences that were overrepresented in the data, the model fails to recognize that a vowel needs to be epenthesized. In addition, even if the model were to have identified that vowel epenthesis is required in the absence of a +CC constraint, the model still cannot choose among the four candidates marked with question marks that were not ruled out by markedness constraints. The CONT-çi constraint does nothing here because there is no vowel in the input for the faithfulness constraint to refer to. There are no constraints against *çe and *çu because there were no words in the training data containing such sequences. What this means is that in C-STAR, the winner is chosen at random when there are multiple candidates that have the same total weight, showing again that a combination of CONT and markedness constraints fail to identify the correct vowel to epenthesize.

Table 6: Evaluation of [çto] 'person' by C-STAR. Wrong winner marked ✗.

[çto]	CONT-to 0.001499	*ça 0.000428	CONT-çi 0.000326	CONT-çt 0.000133	*Total Weight*
✗ /çto/					0
/çato/		-1		-1	-0.000621
? /çito/				-1	-0.000410
? /çuto/				-1	-0.000410
? /çeto/				-1	-0.000410
? /çoto/				-1	-0.000410

For P-STAR, in contrast, the lack of a vowel in the input is not a problem since the positive constraints are only concerned with the output candidates. Table 7 shows how the same word [çto] 'person' from Table 6 was evaluated by P-STAR. Although the faithful candidate is rewarded by the +çt constraint, it still loses to the correct winner /çito/ because the constraint +çi has a higher weight.

Table 7: Evaluation of [çto] 'person' by P-STAR. Clear winner.

[çto]	+to 0.001499	*ça 0.000428	+çi 0.000326	+çt 0.000133	*Total Weight*
/çto/	+1			+1	0.001632
/çato/	+1	-1			0.001071
☞ /çito/	+1		+1		0.001825
/çuto/	+1				0.001499
/çeto/	+1				0.001499
/çoto/	+1				0.001499

5. General discussion and conclusion

This paper proposed a computational model that uses the frequency-driven constraint induction mechanism of STAGE to induce phonotactic constraints for the purposes of learning the phonological process of recovering a reduced high vowel in Japanese. The model was quite successful at learning to recover reduced high vowels, even without access to features that would allow for more general constraint induction. However, implementing the generalization mechanism from STAGE and a feature-level representation to the model could help resolve the issue of /sC/ clusters being consistently epenthesized with /a/ (see Table 3). It is possible that this high frequency of /sa/ is due to the word generator being based on a corpus of newspapers, which has a high occurrence of Sino-Japanese words. Whang (2021) showed that in a corpus of spontaneous Japanese speech, /su/ is in fact the most frequent s-initial CV sequence. Regardless, this frequency discrepancy shows that if learning is strictly input based, the grammar ultimately acquired will change depending on the data the model receives. The model's task, however, is not to recover just any vowel but high vowels. To state differently, there is a rather strict restriction on what CVC sequence an overt CC sequence can correspond to in Japanese—the V is a high vowel (i.e., either /i/ or /u/ and no other vowel). This greatly narrows down the choice of candidates, and since the only high vowel that can follow /s/ in Japanese is /u/, the choice becomes a rather simple one. Because the weight of a general constraint is the sum of all specific constraints supporting it divided by the number of biphones affected by the generalization in STAGE, it can never outrank the highest specific constraint that supports it. So if one assumes that the model induces a general constraint in the form of +{*voiceless obstruent*}{*high vowel*}, the highest weighted constraint supporting it would be +su in the input data used for this study. Although +su is ranked lower than +sa, the general and +su constraints together could result in a gang-up effect that wins over +sa, resulting in the correct epenthesis of /u/ after /s/.

Even if the implementation of a feature-based representation and a generalization mechanism were to fail in correcting the issue with /s/, this is not fatal from a developmental point of view, since infants eventually learn and comprehend words. With the addition of a lexicon, the grammar can keep track of what overt forms correspond to what meaning (Apoussidou 2007) and eventually acquire a paradigm over the lexicon. Such a paradigm can then override the purely phonotactic a-epenthesis after /s/. The same paradigm would presumably also lead to the reinforcement of the other phonotactic epenthesis rules. If it is indeed true that the association of /s/ with the high vowel /u/ is acquired later, it does bring up the question of whether Japanese speakers treat /s/ differently from the rest of the consonants and whether Japanese children ever make the mistake of epenthesizing /a/ instead of /u/ after /s/. Regardless of the specific status of /s/, how features, generalization, and meaning can be implemented into STAR, and what changes the implementation will bring to how the grammar is learned might be an exciting direction to take for future research.

Lastly, although P-STAR simply replaced contiguity constraints with positive constraints while leaving the markedness constraints unchanged, it could just as easily have been a model that induces only positive constraints. When the markedness constraints were removed, the model performed equally well. While this suggests that the positive constraints were doing all the work in the Eval of P-STAR, it is difficult to be sure with the limited test data set. Whether both the positive and markedness constraints are necessary requires further investigation. Also, by redefining the CONT constraints, the model essentially lost faithfulness constraints. This is not to say that faithfulness constraints play no role in the acquisition of a process like Japanese high vowel reduction. Since the positive constraints are supposed to be a mechanism that ties distinct units that cooccur frequently to function as one, it may be that the threshold for their induction should be much higher than an O/E ratio of 2.0. A survey of the +CV constraints that were essential for vowel recovery actually revealed that their O/E ratios were all 9 or higher. The one exception to this was the case of +su, which was relatively much lower at 3.338. Since grammaticalization of separate units requires heavy use, setting a higher threshold does make sense. Doing so would result in a three-tier system, where markedness, faithfulness, and positive constraints can all be induced by the same grammar. A similar approach that incorporates the three types of grammar is explored in Wilson (2021). Since positive constraints alone were enough to successfully epenthesize the correct vowel, such a three-tier model could presumably perform the functions of both STAGE and STAR, where the markedness and faithfulness constraints perform word segmentation, while the positive constraints perform repairs within the segmented words. How such a grammar would work exactly, however, requires further research.

APPENDIX A

Example of generated words				
m oː	m i dʑ i	w a j uu	o k^{i} i r i	r eː tɕ^{u} uː
j u d e	s a i	d o r u d a	ɕ^{i} k a i	m o t o ɕ^{i} ts u
ɕ^{u} u ts u	t e k^{u} u	o k oː	ɕ o p p^{u} u	r i ts u k^{i} i n dʑ i
a dʑ i	t a tɕ^{i} i	m e m m a	r i ts k e	j a s u r u g^{i} i ɕ a

APPENDIX B

Test data			
One target		Two targets	
word	*gloss*	*word*	*gloss*
pi̲kapika	‘to glitter’	pi̲su̲toru	‘pistol’
pi̲ɕari	‘slapping’		
pu̲tomaiɴ	‘ptomaine’	pu̲ku̲puku	‘pudgy’
pu̲tsuɴ	‘snap’		
ki̲ta	‘north’	ki̲tsu̲ku	‘severely’
ki̲ɕi	‘shore’		
ku̲keː	‘rectangle’	ku̲tɕi̲kazu	‘talkativeness’
ku̲ɸuː	‘scheme’		
ɕi̲ta	‘tongue’	ɕi̲tsu̲koi	‘persistent’

ɕi̲seː	‘posture’		
ɕu̲tai	‘subject’	ɕu̲ku̲seː	‘purge’
ɕu̲ɕoː	‘prime minister’		
tsu̲ki	‘moon’	tsu̲tɕi̲kabe	‘mud wall’
tsu̲ta	‘ivy’		
tɕi̲kaɴ	‘molester’	tɕi̲ku̲ɕoː	‘damn it!’
tɕi̲çoː	‘ground surface’		
ɸu̲ku	‘clothes’	ɸu̲ɕi̲kateː	‘motherless family’
ɸu̲ɕi	‘knuckle’		
çi̲to	‘person’	çi̲su̲teriː	‘hysteria’
çi̲ɕa	‘rook’		
su̲ki	‘to like’	su̲ki̲konomu	‘to be fond of’
su̲su	‘soot’		

REFERENCES

ADRIAANS, FRANS and RENÉ KAGER. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62, 311-331.

APOUSSIDOU, DIANA. 2007. *The Learnability of Metrical Phonology*. PhD Dissertation. University of Amsterdam.

BECKMAN, MARY and ATSUKO SHOJI. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41, 61-71.

BOERSMA, PAUL and BRUCE HAYES. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32, 45-86.

BYBEE, JOAN. 2006. From usage to grammar: The mind’s response to repetition. *Language* 82.4, 711-733.

CUTLER, ANNE, TAKASHI OTAKE, and JAMES M. MCQUEEN. 2009. Vowel devoicing and the perception of spoken Japanese words. *Acoustical Society of America* 125.3, 1693-1703.

DUPOUX, EMMANUEL, KAZUHIKO KAKEHI, YUKI HIROSE, CHRISTOPHE PALLIER, and JACQUES MEHLER. 1999. Epenthetic vowels in Japanese: A perceptual illusion?

Journal of Experimental Psychology: Human Perception and Performance 25, 1568-1578.

__________, ERIKA PARLATO, SONIA FROTA, YUKI HIROSE, and SHARON PEPERKAMP. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64, 199-210.

ENDRESS, ANSGAR D. and LUCA L. BONATTI. 2007. Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105, 247-299.

HAYES, BRUCE. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, and K. Wheatley (eds.). *Formalism and Functionalism in Linguistics, Volume 1: General Papers*.

__________. 2004. Phonological acquisition in Optimality Theory: The early stages. In R. Kager, J. Pater, and W. Zonneveld (eds.). *Constraints in Phonological Acquisition*.

__________ and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.3, 379-440.

KAJIKAWA, SACHIYO, LAUREL FAIS, RYOKO MUGITANI, JANET F. WERKER, and SHIGEAKI AMANO. 2006. Cross-language sensitivity to phonotactic patterns in infants. *JASA* 120.4, 2278-2284.

KALASHNIKOVA, MARINA, IRIS-CORINNA SCHWARZ, and DENIS BURNHAM. 2016. OZI: Australian English Communicative Development Inventory. *First Language* 36.4, 407-427.

KIMPER, WENDELL. 2016. Positive constraints and finite goodness in Harmonic Serialism. In J. J. McCarthy and J. Pater (eds.). *Harmonic Grammar and Harmonic Serialism*.

MAEKAWA, KIKUO and HIDEAKI KIKUCHI. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In J. van de Weijer, K. Nanjo and T. Nishihara (eds.). *Voicing in Japanese*.

MUGITANI, RYOKO, LAUREL FAIS, SACHIYO KAJIKAWA, JANET F. WERKER, and SHIGEAKI AMANO. 2007. Age-related changes in sensitivity to native phonotactics in Japanese infants. *Journal of the Acoustical Society of America* 122.3, 1332-1335.

OGASAWARA, NAOMI. 2013. Lexical representation of Japanese high vowel devoicing. *Language and Speech* 56.1, 5-22.

_________________ and NATASHA WARNER. 2009. Processing missing vowels: Allophonic processing in Japanese. *Language and Cognitive Processes* 24.3, 376-411.

PATER, JOE. 2012. Serial harmonic grammar and berber syllabification. In T. Borowsky, S. Kawahara, T. Shinya, and M. Sugahara (eds.). *Prosody Matters: Essays in Honor of Lisa Selkirk* 43-72. London: Equinox Press.

PRINCE, ALAN. 2007. The pursuit of theory. In P. de Lacy (ed.). *Cambridge Handbook of Phonology*. Cambridge University Press.

____________ and PAUL SMOLENSKY. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA, and Oxford, UK: Blackwell.

SHAW, JASON and SHIGETO KAWAHARA. 2018. The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics* 66, 100-119.

SHIBATANI, MASAYOSHI. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

TAMAOKA, KATSUO and SHOGO MAKIOKA. 2004. Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, and Computers* 36.3, 531-547.

TESAR, BRUCE. 1995. *Computational Optimality Theory*. PhD Dissertation. University of Colorado.

____________ and PAUL SMOLENSKY. 1996. *Learnability in Optimality Theory* (long version). Technical Report, Department of Cognitive Science.

_______________________________. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29.2, 229-268.

TSUJI, SHO, KENJYA NISHIKAWA, and REIKO MAZUKA. 2014. Segmental distributions and consonant-vowel association patterns in Japanese infant- and adult-directed speech. *Journal of Child Language* 41.6, 1276-1304.

VÄLIMAA-BLUM, RITTA. 2009. Phonotactic constraints in cognitive phonology. *CORELA* 7.1.

WHANG, JAMES. 2018. Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America* 143.2, 1159-1172.

_____________. 2021. Multiple sources of surprisal affect illusory vowel epenthesis. *Frontiers in Psychology* 12, 677571.

WILSON, COLIN. 2021. Static Harmonic Grammar: Constraint conflict without candidate comparison. In *Proceedings of the 38th West Coast Conference on Formal Linguistics*.

James Whang
Department of Linguistics
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul
08826 Korea
email: jamesw@snu.ac.kr

Kakeru Yazawa
Institutes of Humanities and Social Sciences
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki
305-8571 Japan
email: yazawa.kakeru.gb@u.tsukuba.ac.jp