University of Pennsylvania Working Papers in Linguistics

Volume 27
Issue 1 *Proceedings of the 44th Annual Penn Linguistics Conference*
Article 21

7-9-2021

Learnability of a Phonetically Null Segment

Alexandra Nyman
University of Massachusetts, Amherst

Learnability of a Phonetically Null Segment

Abstract

This paper investigates why two classes of French words, both of which contain words that in isolation are phonetically realized as vowel-initial, vary with respect to elision when pronounced right-adjacent to a definite article: one class does not trigger the deletion of the expendable vowel, while the other class does. It is argued that these two classes differ in that one contains underlyingly vowel-initial words, while the other class consists of words whose underlying representations contain an underspecified consonant segment (termed a ghost consonant; Kiparsky, 2003). The paper also addresses how a learner could posit a ghost consonant in the underlying representation of this second class of words given that the ghost consonant is phonetically null. Through Inconsistency Detection (Tesar, 2004) and Presence Feature setting (Nyman & Tesar, 2019) the learner is able to posit a phonetically null segment, resulting in the resolution of a ranking paradox that would otherwise obtain.

Learnability of a Phonetically Null Segment

Alexandra Nyman*

1 Introduction

This work investigates the learnability of an abstract segment in an Optimality Theoretic (Smolensky and Prince 1993) framework with an Output Driven learner (Tesar 2014), showing how a learner could posit a segment with no phonetic realization. The positing of this segment resolves a ranking paradox that would otherwise arise on the basis of surface forms, and is made possible through Inconsistency Detection (Tesar 2004) and by setting the segment's presence feature (Nyman & Tesar 2019). The case study utilized is prevocalic, expendable vowel deletion in French (also referred to as elision). There are two classes of words which, when pronounced in isolation, are phonetically vowel-initial. One of these classes (the so-called h-aspiré words), when pronounced right-adjacent to the definite article [lə], does not trigger the elision of [ə], while the other class does. It is argued that the words of the h-aspiré class are underlyingly consonant-initial due to the presence of a ghost consonant at their onset. Under this assumption, the learner is able to account for unexpected surface forms.

This paper is organized as follows: Section 2 details prevocalic, expendable vowel deletion in French and then motivates the inclusion of the ghost consonant at the onset of h-aspiré words. Section 3 presents the learning algorithm, Section 4 describes the role of the presence feature in the learning algorithm, and Section 5 provides a brief discussion.

2 Background

2.1 Case Study: H-aspiré

French words have three possible initial segments. They can start with a vowel, a consonant, or a third type of segment referred to as h-aspiré. H-aspiré-initial words are pronounced as though they are vowel-initial in isolation. Though this pronunciation in isolation would suggest that h-aspiré words are in fact vowel-initial, they tend to enforce phonological processes that only occur due to the presence of a consonant-initial word (Boersma 2007). An example being prevocalic, expendable, vowel deletion (also referred to as *elision*), involving the deletion of expendable vowels in prevocalic position to avoid hiatus (as is the case in (1a)). This process does not occur, however, in cases where /lə/ is followed by a consonant-initial word as shown in (1b). The expendable /ə/ of /lə/ is retained in the h-aspiré-initial case, shown in (1c).

(1) Prevocalic, expendable vowel deletion in French
  a. Vowel-initial Environment
     lɔm 'the man'
     /lə+ɔm/ → [lɔm]
  b. Consonant-initial Environment
     ləgaʁsɔ̃ 'the boy'
     /lə+gaʁsɔ̃/ → [ləgaʁsɔ̃]
  c. H-aspiré-initial Environment
     ləazaʁ 'the chance'
     /lə+azaʁ/ → [ləazaʁ]

The surface form in (1c) is unexpected, as it features hiatus. The expected surface form would be *[lazaʁ] as a result of elision.

*Many thanks to the members of SoundWorkshop and the Phonology Reading Group at UMass, Amherst for their feedback and comments. Thanks to Michael Becker, Kaden Holladay, Andrew Lamont, Joe Pater, Jonathan Pesetsky, and Bruce Tesar for extensive discussion and support. All mistakes are my own.

Because of these properties, h-aspiré words have been argued to be underlyingly consonant-initial (e.g. Dell (1973) and Boersma (2007) argue that it is glottal stop-initial). There are other accounts, however, which describe that they have a more underspecified segment initially: Bally (1944) characterizes it as a 'zero consonant', Hyman (1985) argues it is a consonant specified for the feature [+consonantal] alone, Prunet (1986) argues it is a consonant without any features, Piggott (1991) argues it is an underspecified-C, and Smolensky and Goldrick (2016) describe it as a weak-C. This paper adheres to this latter proposal: these words do in fact have an underspecified segment at their onset, namely a ghost consonant.

2.2 The Ghost Consonant

Kiparsky (2003) defines a ghost consonant as "an underlying consonantal element with no segmental melody of its own". Examples of ghost consonants surfacing by undergoing featural changes are provided in Zimmerman (2017, 2018) and Kiparsky (2003). This paper, however, proposes that ghost consonants can surface without any featural changes, resulting in the surfacing of a phonetically null segment. This is represented here as ***C*** to mean 'a consonant with a [+consonantal] feature, but no other featural information'.[2] If we assume a ghost consonant is part of the underlying form of h-aspiré words, the underlying form for 'chance', for example, is /***C***azaʁ/. When proceeding /lə/, the resulting mapping is /lə***C***azaʁ/ → [lə***C***azaʁ]. This update to (1c) is shown below in (2).

(2) Prevocalic, expendable vowel deletion in French
Ghost Consonant-Initial Environment
lə***C***azaʁ 'the chance'
/lə̠+***C***azaʁ/ → [lə***C***azaʁ]

Notice here that the ghost consonant surfaces and, as a consequence, there is no instantiation of hiatus. The h-aspiré word is not phonetically consonant-initial, but it is abstractly, or phonologically, ***C***-initial.

One thing of note is that this proposal differentiates ghost segments from latent segments. Latent segments are segments which only surface in particular contexts to resolve Markedness, but otherwise do not surface. From Tranel's (1995) analysis of the properties of latent segments, latent segments are ones which are not assigned a skeletal slot in their corresponding lexical entry and, therefore, cannot be syllabified unless the context motivates the addition of a skeletal slot for it. This, for example, occurs in some analyses of *liaison*, where a latent final consonant only surfaces when preceding a vowel-initial word. Under the theory mentioned above, latent final consonants are not anchored to a skeletal slot, and therefore do not surface if followed by a consonant. If, however, this latent final consonant is followed by a vowel, a skeletal slot for the consonant is inserted, resulting in its successful surfacing. Ghost consonants, on the other hand, are fixed segments and are assigned a skeletal slot (just like typical, non-latent consonants) and will therefore surface unless otherwise restricted by the grammar. To generate the input-output mappings described above in (1a), (1b) and (2), the following constraints in (3) are provided. These constraints enforce prevocalic, expendable vowel deletion.

(3) *V.V: Assign one violation mark for every instance of two adjacent vowels across a morpheme boundary in the output.
Max-V: Assign one violation mark for every vowel in the input that does not have a corresponding output segment.
Dep-C: Assign one violation mark for every consonant in the output which does not have a corresponding input segment.

The ranking of these constraints is dictated by the avoidance of hiatus on the surface. This is shown below in the derivation of /lə.ɔm/ → [lɔm] where (1a), the faithful candidate, violates *V.V. To

[2]This includes not having a place or a continuant specification, differentiating it from another consonant and /ʔ/.

prevent the surfacing of a vowel-vowel sequence, the grammar can either insert a consonant, (1c), or delete /ə/ (1b). As [lɔm] is the surface form, we can conclude that Dep-C is ranked above Max-V.

/lə̲+ɔm/	*V.V	Dep-C	Max-V
a. lə.ɔm	W		L
→ b. lɔm			*
c. lə.ʔɔm		W	L

Table 1: Schwa deletion before a vowel: /lə̲+ɔm/ → [lɔm].

Candidate (1a)'s violation of *V.V entails a preference for the winner, whereas the violation of Max-V in (1b) incurs a preference for the loser, (1a), as does Dep-C. It is clear, then, that *V.V must be ranked above both Max-V and Dep-C in order to achieve the desired output. Instances where the expendable vowel is retained is shown below in Tables 2 and 3. In Table 3, we observe that the ghost-consonant-initial environment behaves identically to the consonant-initial environment.

/lə̲+gaʁsɔ̃/	*V.V	Dep-C	Max-V
→ a. lə.gaʁsɔ̃			
b. l.gaʁsɔ̃			W

Table 2: Schwa surfaces before a consonant: /lə+gaʁsɔ/ → [lə.gaʁsɔ].

/lə+***C***azaʁ/	*V.V	Dep-C	Max-V
→ a. lə.***C***azaʁ			
b. l.***C***azaʁ			W

Table 3: Schwa surfaces before a ghost consonant: /lə+***C***azaʁ/ → [lə.***C***azaʁ].

Table 3 provides evidence of the ghost consonant: if the ghost consonant is part of the underlying form, then no hiatus violations are incurred on the surface. In section 3, it will become clear that deleting consonants is not an option, motivating the ghost consonant to surface. The faithful candidate, (3a) [lə***C***azaʁ] is the winner, just like the faithful candidate in the consonant-initial environment, (2a). Though the inclusion of the ghost consonant can explain why [ə] is retained, the h-aspiré word remains phonetically identical to the vowel-initial word. A question of learnability is raised then: if h-aspiré words and vowel-initial words are phonetically identical, how do learners differentiate them? According to Tesar (2014), if the learner is output-driven, the learner observes a surface form and maps this surface form directly to an identical underlying form. Through the processes of Non-Phonotactic learning and Inconsistency Detection, the learner is able to construct the correct underlying form based on the observed surface form. As will be discussed in Section 3.3, the learner posits a ghost consonant to phonologically differentiate the h-aspiré word from vowel-initial words. This leads us into learning, described in the following sections.

3 Learning

Now we move on to the core question: how do learners posit abstract structure? This requires proper motivation given the learning data. The learnability algorithm[3] presented in this work is dependent on paradigmatic evidence of alternation where a learner observes an unexpected surface structure and concludes that their current grammar is lacking. Importantly, this restricts the positing space from a learning standpoint: the learner only posits a ghost consonant in the underlying form when there is observable evidence to motivate it.[4] The learner begins with a set of information: We assume the learner has access to a non-exhaustive morpheme inventory, Table 4, and a constraint set from the learning data, (4); # indicates a morpheme boundary.

Morpheme	Underlying Form	Word	Candidate
m1	/lə/	m1m2	/lə#ɔʁɑʒ/ → [lɔʁɑʒ]
m2	/ɔʁɑʒ/	m3m4	/kɜl#gaʁsɔ̃/ → [kɜlgaʁsɔ̃]
m3	/kɜl/		
m4	/gaʁsɔ̃/		

Table 4: The learner's morpheme inventory.

(4) Ons: Assign one violation mark for every syllable without an onset.
*V.V: Assign one violation mark for every instance of two adjacent vowels across a morpheme boundary in the output.
*CC: Assign one violation mark for every instance of two adjacent consonants in the output
Max-C: Assign one violation mark for every consonant in the input that does not have a corresponding output segment.
Dep-C: Assign one violation mark for every consonant in the output that does not have a corresponding input segment.
Max-V: Assign one violation mark for every vowel in the input that does not have a corresponding output segment.
Dep-V: Assign one violation mark for every vowel in the output that does not have a corresponding input segment.

With access to this set of constraints and a morpheme inventory consisting of underlying forms and input-output mappings, the learner begins to gather phonotactic ranking information.

3.1 Phonotactic Ranking Information: Part I

The learner begins by collecting Phonotactic ranking information (Tesar and Prince 2003), a set of fully-faithful candidates paired with informative losers. Given the constraints, the learner compiles phonotactic support with winners being the observed form, paired with losers that provide ranking information. Here, the observed forms that have informative losers are: [ɔʁɑʒ] 'orange', and [kɜlgaʁsɔ̃] 'what boy'.

[3]For a fully comprehensive overview of this learning algorithm and its corresponding pseudo-code, see Nyman & Tesar (2019).
[4]Preventing the possibility of unmotivated ghost consonant insertion.

Input	W~L	Ons	*V.V	*CC	Max-C	Max-V	Dep-C	Dep-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W

Table 5: Phonotactic ranking information, Part I.

The informative losers above show that onsetless syllables are allowed over violating Max-V or Dep-C, and consonant clusters are allowed over violating Max-C or Dep-V. This leaves the learner with an initial ranking of the constraints. It is not until the learner encounters alternation that the ranking will need to be updated.

3.2 An Alternation: Prevocalic, Expendable Vowel Deletion

After learning from fully faithful mappings, the learner considers unfaithful mappings like: /lə#ɔʁɑʒ/ → [lɔʁɑʒ] 'the orange'. The winner/loser pairs for this candidate as compared to the current, Phonotactic ranking information is provided in Table 6, under the bolded line.

Input	W~L	Ons	*V.V	*CC	Max-C	Max-V	Dep-C	Dep-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləɔʁɑʒ	W	W			L		
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləʔɔʁɑʒ					L	W	

Table 6: Non-Phonotactic ranking information: /lə#ɔʁɑʒ/ → [lɔʁɑʒ].

The informative losers in this case, [ləɔʁɑʒ] and [ləʔɔʁɑʒ], show that the grammar prefers the deletion of a vowel over the insertion of a consonant. Because this alternation provides the learner with new information regarding the preferences of their grammar, the learner gains Non-Phonotactic ranking information (Tesar 2006), which is then imported into the phonotactic ranking information.

3.3 Phonotactic Ranking Information: Part II

The candidates shown above in Table 6 have provided the learner with relevant information regarding the relative ranking of Dep-C and Max-V: the grammar prefers Dep-C over Max-V and will delete a vowel over inserting a consonant. Therefore, the learner adds these non-phonotactic candidates to the phonotactic ranking information (shown in the final two rows of Table 7 below).[5]

After importing the non-phonotactic ranking information, the learner proceeds with Biased Constraint Demotion (BCD) (Tesar 2002), with a Markedness Bias, to determine the ranking given the current language data. The ranking after BCD is shown below in (5).

Input	W~L	Ons	*V.V	*CC	Max-C	Max-V	Dep-C	Dep-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləɔʁɑʒ	W	W			L		

[5]See Tesar (2014) for more information regarding the process of updating phonotactic ranking information.

lə#ɔʁaʒ	lɔʁaʒ ~ ləʔɔʁaʒ					L	W	

Table 7: Phonotactic ranking information: Part II.

(5) *V.V >> {Dep-C, Dep-V, Max-C} >> *CC >> Max-V >> Ons.

Because Dep-C and Max-C are so highly ranked, the learner is left with few options to account for the unexpected surface form, [ləazaʁ]. It is at this point that the learner resorts to positing a ghost consonant. Since the insertion of a ghost consonant will not help, nor will a ghost consonant in the underlying form, set to be deleted, help, the learner determines that the ghost consonant must be in the surface form, as well as the underlying form (see Table 9). After the ranking is ascertained, the learner can test multiple options. Table 8 illustrates that the learner's grammar will predict the wrong output if the ghost consonant is not in the underlying form:

/lə#azaʁ/	*V.V	Dep-C	Dep-V	Max-C	*CC	Max-V	Ons
a. [ləazaʁ]	*!						*
b. [lə***C***azaʁ]		*!					
*→c. [lazaʁ]						*	
d. [l***C***azaʁ]		*!			*	*	

Table 8: Incorrect winner: /lə#azaʁ/ → *[lazaʁ].

To achieve the target output, the learner posits a ghost consonant in the underlying form. This is shown in Table 9:

/lə#***C***azaʁ/	*V.V	Dep-C	Dep-V	Max-C	*CC	Max-V	Ons
→[lə***C***azaʁ]							
[ləazaʁ]	*!			*			*
[lazaʁ]				*!		*	
[l***C***azaʁ]					*!	*	

Table 9: Correct winner: /lə#***C***azaʁ/ → [lə***C***azaʁ].

Now that the learner has confirmed that this ranking generates the correct surface form after including the ghost consonant in the underlying form, the learner determines that the ghost consonant must be in both the input and output. The learner now continues to presence feature setting, which will secure the presence of the ghost consonant in the underlying form.

4 Setting the Presence Feature

The presence feature is a learning tool intended to indicate the presence or absence of an underlying segment (Nyman & Tesar 2019). It is not a phonological feature, but a learning feature. One consequence of having the property of a learning feature, as opposed to a phonological feature, is that it is strictly used during learning and, therefore, is removed after learning is complete. Its sole purpose is to simplify learning and resolve ambiguity. The presence feature was first introduced within basic CV syllable theory (Clements & Keyser 1983) to learn insertion and deletion. It is now being applied to concrete language data in order to learn abstract surface segments. Table 10 describes the ways in which the learner represents presence and absence.

Symbol	Meaning
+	Segment must be in the input
–	Segment cannot be in the input
?	Unset

Table 10: Notations for the possible states of a presence feature.

To determine the presence of a segment in the underlying form, the learner utilizes a learning process called Inconsistency Detection (ID) (Tesar 2004). ID is a process by which the learner tests a possible candidate against its current phonotactic ranking information to see if that candidate will be generated by their grammar. If the grammar can generate the desired output, then it is consistent and the learner continues. If the candidate is inconsistent, the learner's current grammar is lacking and the learner must re-evaluate. This is detailed further, with examples, in the next section. During presence feature setting, a segment is set to (+presence) underlyingly, for example, if there is paradigmatic evidence through ID that the segment in question must be present underlyingly to achieve the desired output. The testing candidate includes the temporary setting of the presence feature for all posited segments of the morpheme. The target segment is set to the opposite of its surface realization: if it is on the surface, it is tested as (-presence) underlyingly; if not, then it is set to (+presence) underlyingly. All other segments are set to (+presence) to match the surface realization.[6]

4.1 Setting the Presence Feature of the Ghost Consonant

Once phonotactic learning is complete, the learner has access to the morphemic constituency of each word (which includes the morphemic affiliation of each segment of each output). Whenever the learner encounters a new morpheme, it creates an entry in its lexicon for that morpheme with one underlying segment for each segment of the observed output (with all presence features unset). The learner then constructs an input-output correspondence based on the information given by the morphemic affiliations (see Table 11).

Given the motivation behind asserting a ghost consonant, the learner posits the ghost consonant in the output and includes it in the input entry for m5. This entry includes the presence features of each posited segment, initially unset.

Morpheme	Underlying Form
m5	/(?,***C***)(?,a)(?,z)(?,a)(?,ʁ)/

Table 11: Example inventory after entry creation for morpheme m5.

The learner now tests the presence feature for the ghost consonant. Because the ghost consonant is posited as present in the output, the test value of the feature is (–presence). All other unset features are temporarily assigned the value matching its surface realization, (+presence). The candidate to be tested for consistency is shown below. The segment which is the target for feature setting is underlined.

(6) /(<u>**-,*C***</u>)(+,a)(+,z)(+,a)(+,r)/ = /azaʁ/ → [***C***azaʁ]

With this testing candidate as the input, the learner selects an informative loser, [azaʁ] to be paired with the desired winner, [***C***azaʁ]; these form the winner-loser pair shown in the bottom row of (12).

Input	W~L	Ons	*V.V	*CC	Max-C	Max-V	Dep-C	Dep-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləɔʁɑʒ	W	W			L		
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləʔɔʁɑʒ					L	W	

[6]The opposite is true of (-): a segment is set to (-) when the segment cannot be there underlyingly (cases of insertion). (?) when the presence of a segment is unknown and/or cannot be set.

azaʁ	***C***azaʁ ~ azaʁ	W					L	

Table 12: Candidate /azaʁ/ → [***C***azaʁ] as compared to learner's phonotactic ranking information

The inconsistency between the testing candidate and the learner's current grammar is indicated by the shaded cells. The inconsistent candidate resulted from assigning the value (–presence) to the underlying presence feature of /***C***/. This inconsistency justifies setting the feature to (+presence) in the morpheme inventory, resulting in the entry update for m5, shown in (13):

Morpheme	Underlying Form
m5	/(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/

Table 13: Example inventory after feature setting for /***C***/.

The learner must now determine whether /ə/ is present or absent in the underlying form of m1: /lə/. As the learner has encountered m1 and m3 before, the morphemic identity of each morpheme in [lɔʁɑʒ] is already known. However, the learner does not know whether /ə/ is present in the underlying form of m1, and is deleted when m1 is adjacent to m3, or if it is absent in the underlying form of m1. To determine whether or not /ə/ is present or absent, the learner targets this segment for presence feature setting. The testing candidate is first m1m3 (the [lɔʁɑʒ] case).

Morpheme	Underlying Form
m5 m1m3	/(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/ /(?,l)(?,ə)#(?,ɔ)(?,ʁ)(?,ɑ)(?,ʒ)/

Table 14: Example inventory after entry creation for word m1m3.

Because /ə/ is not present in the output, the test value of the feature is (+presence) and all other unset features are temporarily assigned the value matching its surface realization, (+presence). The candidate to be tested for consistency, therefore, is shown below:

(7) /(+,l)(+,ə)#(+,ɔ)(+,ʁ)(+,ɑ)(+,ʒ)/= /lə#ɔʁɑʒ/ → [lɔʁɑʒ]

As we saw in non-phonotactic learning, whichever informative loser the learner picks with this underlying form will derive consistency so /ə/ will remain unset. The learner chooses [ləʔɔʁɑʒ] as the informative loser, shown in the last row of Table 15 below.

The candidate is consistent, as expected, so the presence feature for /ə/ remains unset. At this point, the learner does not know whether /ə/ is present underlyingly since it has not surfaced in any forms it has seen during phonotactic learning. However, the learner now tests the presence of /ə/ in another context, namely, m1m5 (the [lə***C***azaʁ] case). See Table 16 below.

Input	W~L	ONS	*V.V	*CC	MAX-C	MAX-V	DEP-C	DEP-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləɔʁɑʒ	W	W			L		
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləʔɔʁɑʒ					L	W	
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləʔɔʁɑʒ					L	W	

Table 15: Candidate /lə#ɔʁɑʒ/ → [lɔʁɑʒ] is consistent.

Morpheme	Underlying Form
m5	/(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/
m1m3	/(?,l)(?,ə)#(?,ɔ)(?,ʁ)(?,ɑ)(?,ʒ)/
m1m5	/(?,l)(?,ə)#(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/

Table 16: Example inventory after entry creation for word m1m5.

Because /ə/ surfaces in m1m5, the test value of the feature is (–presence). All other unset features are temporarily assigned the value matching its surface realization, (+presence). The candidate to be tested for consistency is shown below:

(8) /(+,l)(-,ə)#(+,*C*)(+,a)(+,z)(+,a)(+,r)/ = /l***C***azaʁ/ → [lə***C***azaʁ]

The learner selects, as an informative loser, [l***C***azaʁ], and forms the winner-loser pair shown in the bottom row of Table 17:

Input	W~L	Ons	*V.V	*CC	Max-C	Max-V	Dep-C	Dep-V
ɔʁɑʒ	ɔʁɑʒ ~ ʔɔʁɑʒ	L					W	
ɔʁɑʒ	ɔʁɑʒ ~ ʁɑʒ	L				W		
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜgaʁsɔ̃			L	W			
kɜl#gaʁsɔ̃	kɜlgaʁsɔ̃ ~ kɜləgaʁsɔ̃			L				W
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləɔʁɑʒ	W	W			L		
lə#ɔʁɑʒ	lɔʁɑʒ ~ ləʔɔʁɑʒ					L	W	
l*C*azaʁ	lə***C***azaʁ ~ l***C***azaʁ			W				L

Table 17: The candidate /l***C***azaʁ/ → [lə***C***azaʁ] is inconsistent.

The inconsistent candidate resulted from assigning the value (–presence) to /ə/, and the inconsistency justifies setting the feature to (+presence) in the morpheme inventory, resulting in the entry shown in (18):

Morpheme	Underlying Form
m5	/(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/
m1m3	/(?,l)(+,ə)#(?,ɔ)(?,ʁ)(?,ɑ)(?,ʒ)/
m1m5	/(?,l)(+,ə)#(+,***C***)(?,a)(?,z)(?,a)(?,ʁ)/

Table 18: Presence feature setting for /ə/.

/ə/ is set to (+presence) for all inventory entries that include m1. The learner now knows that it is present underlying for any instance where m1 is part of the underlying form. This means the learner not only knows that it is present underlyingly when it surfaces, it is also present underlyingly but deleted in the case of m1m3. In general, the learner would continue to process until it is forced to halt. It is fully possible for a presence feature to remain unset: if there is no paradigmatic evidence providing motivation to set the presence feature, the learner will reach consistency every time.

5 Discussion

The successful learning of /ə/ deletion in the case of [lɔʁɑʒ] was due to the setting of that segment's presence feature. Similarly, the learner was able to successfully set the presence feature for the ghost consonant as present underlyingly, substantiating the learner's hypothesis that the ghost consonant was part of the underlying form. This success first required the positing of the ghost consonant, which was motivated by the ranking paradox observed when two phonetically identical words enforced different surface structures. Though the learning algorithm presented here differs greatly from previously proposed, and quite robust, learning models of h-aspiré processes, (e.g. Smolensky and Goldrick 2016), this paper illustrated that an output-driven learner, equipped with presence feature setting, can learn that a phonetically null segment is part of an underlying form.

References

Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117:1989-2054.

Clements, George, and Samuel Jay Keyser. 1983. *Cv Phonology. A Generative Theory of the Syllable*. Linguistic Inquiry Monographs 9:1-191.

Dell, François. 1973. Les règles et les sons. *Introduction to Generative Phonology*:360-365.

Hyman, Larry. 1985. *A Theory of Phonological Weight*. Dordrecht: Foris.

Kiparsky, Paul. 2003. Finnish noun inflection. In *Generative Approaches to Finnic and Saami Linguistics*. ed. D. Nelson and S. Manninen, 109-161. Stanford, CA: CSLI Publications.

Nyman, Alexandra, and Bruce Tesar. 2019. Determining underlying presence in the learning of grammars that allow insertion and deletion. *Gloss*a 4:1-41.

Piggott, Glyne. 1991. Empty onsets: evidence for the skeleton in prosodic phonology. *McGill Working Papers in Linguistics/Cahiers linguistiques de McGill* 7:41-71.

Prunet, Jean-François, 1986. Spreading and Locality Domains in Phonology. Doctoral dissertation, McGill University.

Smolensky, Paul, and Alan Prince. 1993. Optimality Theory: Constraint interaction in generative grammar. In *Optimality Theory in Phonology: A Reader*. ed. J. McCarthy, 1-71. Cambridge, MA, Oxford: Blackwell.

Smolensky, Paul, and Matthew Goldrick. 2016. Gradient symbolic representations in grammar: The case of French liaison. Rutgers Optimality Archive:1552.

Tesar, Bruce. 2002. Enforcing grammatical restrictiveness can help resolve structural ambiguity. *Proceedings of the West Coast Conference on Formal Linguistics 21*:443-456.

Tesar, Bruce. 2004. Using Inconsistency Detection to Overcome Structural Ambiguity. *Linguistic Inquiry* 35: 219-253.

Tesar, Bruce. 2014. *Output-driven Phonology: Theory and Learning*. Cambridge University Press.

Tranel, Bernard. 1995. Current issues in French phonology. In *The Handbook of Phonological Theory*, ed. J.A. Goldsmith, 798–816. Cambridge, MA: Blackwell.

Tranel, Bernard. 1996. French liaison and elision revisited: A unified account within Optimality Theory. *Aspects of Romance linguistics*: 433-455.

Zimmermann, Eva. 2017. Gradient symbolic representations in the output: A typology of lexical exceptions. Paper presented at NELS 48, University of Iceland, Reykjavik.

Zimmermann, Eva. 2018. Gradient Symbolic Representations and the Typology of Ghost Segments: An Argument from Gradient Markedness. Paper presented at the 2018 Annual Meeting on Phonology, UC San Diego.

Department of Linguistics
University of Massachusetts, Amherst
Amherst, MA 01002
anyman@umass.edu