

# *Emerging grammars in contemporary Yoruba phonology*

TAOFEEQ ADEBAYO 

*California State University, San Bernardino, California, USA*

[taofeeq.adebayo@csusb.edu](mailto:taofeeq.adebayo@csusb.edu)

---

## *Abstract*

This article provides a description and an Optimality Theory (OT) analysis of contact-induced changes and variation in contemporary Yoruba syllable structure. The article claims that a major diachronic change has occurred in the syllable structure of Yoruba phonology due to its continued contact with English, resulting in the invention, preservation, and hypercorrection of clusters and codas. I characterize this change in terms of OT constraint re-ranking (Miglio and Moren 2003) and assess the resulting synchronic variation against the indexed constraint approach of Itô and Mester (1995a, b, 1999), the ranked-winners approach of Coetzee (2004), the partial-order co-phonology of Anttila (1997), and the Maximum Entropy (MaxEnt) model of Goldwater and Johnson (2003). I show that none of these approaches is able to account independently for the categorical, gradient, and lexically conditioned variation that characterize the contemporary Yoruba syllable structure, but rather that a MaxEnt model augmented with lexical indexation is the most economical model that fits the Yoruba data accurately.

**Keywords:** Yoruba, clusters, codas, diachronic change, variation

## *Résumé*

Cet article fournit une description et une analyse dans la théorie de l'optimalité (OT) des changements et de la variation induite par le contact dans la structure syllabique du yoruba contemporain. L'article affirme qu'un changement diachronique majeur s'est produit dans la structure syllabique de la phonologie yoruba en raison de son contact continu avec l'anglais, ce qui entraîne l'invention, la préservation et l'hypercorrection des groupes consonantiques et des

---

This article is dedicated to Olanike Ola Orie, who passed away on 9/11/2021. I am very grateful for her comments and suggestions on all the previous versions of this article. I am also grateful to Akinlabi Akinbiyi and John Singler, both of whom offered useful suggestions on previous versions of the article. Thanks also to the three anonymous reviewers for their suggestions.

codas. Je caractérise ce changement en termes de reclassement des contraintes OT (Miglio et Moren 2003) et j'évalue la variation synchronique résultante d'après l'approche des contraintes indexées d'Itô et Mester (1995a, b, 1999), l'approche des gagnants classés de Coetzee (2004), la co-phonologie d'ordre partiel d'Anttila (1997) et le modèle d'entropie maximale (MaxEnt) de Goldwater et Johnson (2003). Je montre qu'aucune de ces approches n'est capable de rendre compte de manière indépendante de la variation catégorielle, graduelle et lexicalement conditionnée qui caractérisent la structure syllabique du yoruba contemporain, mais que le modèle le plus économique et qui correspond le mieux aux données yoruba est un modèle MaxEnt augmenté de l'indexation lexicale.

**Mots-clés:** yoruba, groupes consonantiques, codas, changement diachronique, variation

## 1. INTRODUCTION

This article claims that Yoruba has changed from a language that makes categorical choices against consonant clusters and codas to a language that variably permits them through invention, preservation, and hypercorrection<sup>1</sup>. This diachronic change is conceptualized in terms of Optimality Theory (OT) constraint re-ranking in line with the assumptions of Miglio and Moren (2003). The resulting variation is assessed against existing OT approaches to synchronic variation and lexical stratification, including the ranked-winners approach of Coetzee (2004), the indexed constraint approach of Itô and Mester (1995a, b, 1999), the partial-order co-phonology of Anttila (1997), and the Maximum Entropy (MaxEnt) model of Goldwater and Johnson (2003). None of these approaches is able to account for the variation without further mechanisms. However, it is shown that a MaxEnt model augmented with lexical indexation provides the most successful account of the Yoruba data.

I identify some major contributions of this article to linguistic theory. First, I show how different theoretical approaches to a particular phenomenon can be harmonized in the face of new data. Second, I reveal a new kind of lexical stratification and synchronic variation that may inform linguistic theory. More specifically, I show how a core-periphery structure of the lexicon can be organized differently from the patterns documented in Hsu and Jesney (2018). Third, I introduce conflation tableaux that represent several tableaux simultaneously. Lastly, I show how change in progress can be modeled in a linguistic theory predicting synchronic variation only if it is, indeed, a matter of change in progress rather than stable variation.

Consensus on Yoruba phonology is that it disallows both consonant clusters and codas (Ola 1995, Akinlabi 2004, etc.). These facts shape loanword outputs in

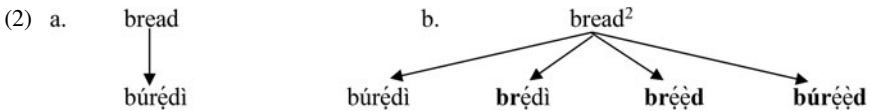
---

<sup>1</sup>Abbreviations used: CIS: Cluster Hypercorrection Periphery Stratum; CMS: Church Missionary Society; CoS: Coda Hypercorrection Periphery Stratum; CS: Core Stratum; CV: Consonant, Vowel; GLA: Gradual Learning Algorithm; LI : lexical item; MaxEnt: Maximum Entropy; OT: Optimality Theory. Syllables are separated by ‘.’, and standard Yoruba orthography is used for Yoruba words and adapted loanwords throughout this article:  $\text{ɕ} = \text{ʃ}$ ,  $\text{ɛ} = [\text{ɛ}]$ ,  $\text{ɔ} = [\text{ɔ}]$ , ‘ = high tone, ` = low tone, unmarked for tone = mid tone. /kp/ and /gb/ are single phonemes not clusters, and the letter p in a Yoruba word can represent either /p/ or /kp/.

Yoruba, so that if a loanword has clusters, the clusters are resolved by epenthetic vowels (1a–d), and if a loanword contains a coda, the coda is resolved by deletion, as in (1c), or epenthetic vowel insertion, as in (1e).

(1)	<b>English</b>	<b>Output in Yoruba</b>	<b>Gloss</b>
a.	bread	búrèdì	bread
b.	blade	bílèèdì	blade
c.	school	sùkùù	school
	<b>Arabic</b>	<b>Output in Yoruba</b>	<b>Gloss</b>
d.	fitna	fitnà	temptation
e.	naas	nèésì	mankind

Most loans in Yoruba come from English and Arabic, but most recent loans are from English. Most Arabic loans have diffused into Yoruba to the extent that most Yoruba speakers do not even know their origin (see Badmus 2020: 386 for example). One such example is *sánmà*, “sky”, adapted from the Arabic word *samah*. However, given the increase in the number of English–Yoruba bilingual speakers, coupled with the prestige of English among the Yoruba people in Nigeria, these adaptation strategies have dramatically changed in the synchronic grammar of Yoruba speakers, whether bilingual or monolingual, and whether they live in a rural or an urban area. Most English loans in Yoruba today are adapted with clusters (observed in Ufomata 1991 and in Adeniyi 2015) and codas (also recognized in Ufomata 1991), both in writing and in speech. In *Aláròyè*, a popular Yoruba newspaper, for example, we find *mínístà* ‘minister’, *fráide* ‘Friday’, etc. *Aláròyè* (Alaroye TV 2021) uses the title ‘Tinubu n sun l’owò, ẹ ma **DISTOQBU** re’ where the cluster in *distoqbu*, ‘disturb’, is retained while the coda is resolved. Yoruba movies, radio, and television programs are also permeated with English loanwords adapted with their clusters and codas intact. These new adaptation strategies have, however, given rise to contact-induced variation so that for example, rather than have a situation as in (2a), what we now have is a situation where (2b) holds. A pattern similar to (2b) where a single loan form has more than one variant is also found in Kang et al. (2008) for Korean.



I analyze these emerging phonological phenomena in Yoruba as evidence for a major diachronic change in the grammar of the language. To this end, I present evidence

<sup>2</sup>A reviewer asks whether any given Yoruba speaker will accept/use all the four forms. I offer the following observations: forms like *búrèdì*, *brèdì*, and *brèèd* are more frequent than *búrèèd*; based on this, we might conclude that the former forms are better established than the latter, and that consequently any given Yoruba person is likely to accept any of the three forms but that the *búrèèd* form might be restricted.

from neologisms (such as *fléjìḡ* ‘enjoy’ and *wòós* ‘hey’), which are formed with either clusters or codas. I also present evidence from a cluster hypercorrection phenomenon in which both native Yoruba words (3a) and loanwords (3b) now undergo vowel deletion in an attempt to create consonant clusters where they did not originally exist. Further evidence is adduced from coda hypercorrection, where vowel deletion results in word-final codas (3c).

- |     |                  |                        |                |
|-----|------------------|------------------------|----------------|
| (3) | a. bẹ súkẹ →     | bésékẹ                 | ‘soybean cake’ |
|     | b. paracetamol → | parastamòḡ ~ prastamòḡ |                |
|     | c. bẹ̀è ni →     | bẹ̀èn                  | ‘yes’          |

In order to motivate these diachronic changes, I propose the following periods in the development of Yoruba phonology in accordance with its contact with English. According to Taiwo (2009), English first arrived in Nigeria in the late sixteenth century when British merchants and Christian missionaries settled in the coastal town of Badagry, near Lagos. Since written documents in Yoruba do not emerge until much later, little or nothing is known about the Yoruba of that time period. I will refer to the Yoruba spoken at that period as **Contact Yoruba**, and the period before the arrival of the British as **Pre-Contact Yoruba**. The first recorded collection of Yoruba words was written by Bowdich in 1817 and published in 1819 in *Mission from Cape Coast to Ashantee* (Bamgbose 1982). Other written documents followed, and these include the vocabulary collections by Hannah Kilham in 1828, Clapperton in 1829, and sample-sentence and vocabulary collections by John Raban in 1830, 1831, and 1832 (Bamgbose 1982). Major work on Yoruba grammar did not emerge until the 1840s when Bishop Samuel Ajayi Crowther published his *Vocabulary of the Yoruba language* and began the translation of the Bible into Yoruba. I will refer to the Yoruba represented in the collections by Bowdich in 1817 through to those by Raban in 1832 as **Pre-Crowther Yoruba**, and the Yoruba spoken beginning with the work of Ajayi Crowther and the Church Missionary Society (CMS) in Lagos till the end of the nineteenth century as **Crowther Yoruba**. The Yoruba spoken from the beginning of the twentieth century until the present is referred to as **Contemporary Yoruba**.

Since we do not have concrete information about Pre-Contact Yoruba or Contact Yoruba, I will not discuss them further. The Yoruba recorded in Pre-Crowther Yoruba is one characterized by impermissibility of clusters and codas. Although the grammars presented in Crowther Yoruba mainly portray Yoruba as a language without clusters or codas, the first recorded instance of cluster retention is at the end of the nineteenth century. In the Yoruba Bible that emerged during that period, the word ‘Christ’ was adapted as *kristi*, where the cluster is retained and the coda is resolved. The first recorded cluster retention in Contemporary Yoruba can be found in *A Dictionary of Yoruba Language* published by the Church Missionary Society in 1913 where the word ‘Christ’ was also adapted as *krístì* and the word ‘apostle’ was adapted as *àpóstìlì* (CMS 1913). The first recorded instance of cluster and coda retention in spoken speech did not emerge until Ufomata (1991). Adeniyi (2015) is a more recent attestation of cluster retention in spoken Contemporary Yoruba. It follows then that we can recognize two major periods in

the development of Yoruba phonology: Period 1, where clusters and codas are not permitted in any word (Pre-Crowther Yoruba) and Period 2, where codas and clusters are permitted in some words (Crowther Yoruba and Contemporary Yoruba). Period 3 is a period where the change (in progress) from Period 2 has gone to completion or continues to mirror Period 2. This will be schematized in [Figure 2](#), found in section 4.1.

My focus in this article is on Periods 1 and 2. I briefly look at the transition between the two periods and analyze the patterns found in Period 2 as synchronic variation. I claim that the transition from Period 1 to Period 2 constitutes a major diachronic change in the grammar of the language, a movement from total impermissibility of clusters and codas to permissibility of clusters and codas in some words. The question of whether this synchronic variation is a stable variation that will persist, or else a change in progress that will lead to eventual completed change (Period 3), I leave to future research.

The data presented in this article are based on more than 120 hours of sociolinguistic interviews, 18 months' worth of participant observation, and grammaticality judgments in Oyo, Lagos, and Ogun states (Nigeria). Details on data collection will be found in Appendix A. The data are presented from the viewpoint of Optimality Theory (Prince and Smolensky 2004 [1993]), where the grammar of a language is expressed in terms of constraint ranking. According to this theory, languages have the same constraints, but differ in how the constraints are ranked (McCarthy 2007). I take British English to be the primary source of all English loans described throughout this article. One may ask why Nigerian English is not the main source of these loans. Two things would make this problematic. First, the English variety named Nigerian English is not homogenous. In fact, it is a collection of several Englishes found in Nigeria, such as Hausa English, Igbo English, and Yoruba English (Gut 2008: 38). One might then ask: Why isn't Yoruba English the main source of the loans? The reason is that it is a variety of English that is conditioned by the Yoruba grammar in the first place. This is not to say that there are no loans (such as *nabteb* 'NABTEB', *k-leg* 'knock knee', etc.) which originate from the collection tagged 'Nigerian English'. Neither can it be denied that Nigerian English and Nigerian Pidgin English may play a role in the ongoing change in the Yoruba syllable structure; but these considerations lie beyond the scope of this article.

The rest of this article is organized as follows. Section 2 describes how consonant clusters and codas are resolved in traditional (Pre-Crowther) Yoruba phonology. In section 3, I outline contact-induced patterns that can now be found in the syllable structure of Contemporary Yoruba phonology. Section 4 presents diachronic and synchronic OT analyses. Section 5 outlines the major predictions of the analyses. Section 6 provides a summary of the major results of the article.

## 2. SYLLABLE STRUCTURE IN PRE-CROWTHER YORUBA PHONOLOGY

According to Ola (1995) and Akinlabi (2004), Yoruba has only two types of syllables: a syllable that contains only a vowel (the V-type) and one that contains a

consonant and a vowel (the CV type). Bilabial and alveolar nasals can be moraic, and so can stand as a syllable on their own, bearing at least a tone. They are categorized under the V type of syllable structure. In the examples in (4), the segments that constitute the respective Yoruba syllable types are in boldface.

(4)	a. alé	[a. <b>lɛ̃</b> ]	consonant–vowel	CV	“night”
	b. owó	[ <b>o</b> .wó]	vowel	V	“money”
	c. olóńgbò	[o.ló. <b>ń</b> .gbò]	nasal	V	“cat”
	d. òrombò	[ò.ro. <b>m</b> .bò]	nasal	V	“lime”

Coding this fact about Yoruba syllable structure into an Optimality-theoretic model follows straightforwardly. Based on (4), we can state the following generalizations about the Pre-Crowther Yoruba phonology: a) Consonant clusters are forbidden; and b) codas are forbidden. The relevant OT well-formedness constraints that ensure that phonological outputs in Pre-Crowther Yoruba do not appear with a cluster or with a coda are \*COMPLEX and NoCODA respectively (5a–b). I follow Itô and Mester (1999) and abbreviate the Input/ Output faithfulness constraints DEP and MAX as FAITH in (5c), since ranking among faithfulness constraints is not essential to the discussions that follow.

- (5)
- a. \*COMPLEX: Complex onset and complex codas are forbidden (McCarthy 2008, originally formulated in Prince and Smolensky 2004 [1993]), whether tautosyllabic or heterosyllabic.
  - b. NoCODA: Syllables must not have codas (Tesar 2004)<sup>3</sup>.
  - c. FAITH: (i) MAX: Every segment of the input has a correspondent in the output: no phonological deletion (McCarthy and Prince 1995, 1999).  
(ii) DEP: Every segment of the output has a correspondent in the input: no phonological epenthesis (McCarthy and Prince 1995, 1999).

In Pre-Crowther Yoruba phonology, the ranking argument in (6) produces the optimal candidates. The tableau in (6) shows that both native words (6.I) and loanwords from Arabic (6.II) and English (6.III) abide by the two high-ranking constraints \*COMPLEX and NoCODA. This suggests that in Pre-Crowther Yoruba phonology, these markedness constraints determine the shape of both native and loan outputs.

<sup>3</sup>In this article, NoCODA is understood as NoCODA/Wd (Flack 2009: 275), defined as ‘word-final syllables must not have codas’. I restrict the analysis to word-final codas in order to focus on clear cases of coda realizations. This is also why \*COMPLEX targets both tautosyllabic and heterosyllabic consonant sequences. Pre-Crowther Yoruba did not permit any clusters or codas. In order to address whether Yoruba now permits them, we can look at word-final codas and consonant sequences. How word-medial codas arise from heterosyllabic clusters is an independent question that I address in Adebayo (2022).

(6) Pre-Crowther Yoruba ranking argument: \*COMPLEX, NoCODA >> FAITH

	*COMPLEX	NoCODA	FAITH
<b>I. /oní bàtà/ ‘shoe owner’</b>			
☞ (a) [o.ní.bà.tà]			
(b) [ó.ní.bàt]		*!	*MAX
(c) [o.nbá.tà]	*!		*MAX
<b>II. /fitna/ ‘temptation’</b>			
☞ (a) [fi.ti.na]			*DEP
(b) [fi.tna]	*!		
(c) [fi.ti.in]		*!	
<b>III. /brɔd/ ‘bread’</b>			
☞ (a) [bú.ré.dì]			**DEP
(b) [bré.dì]	*!		*DEP
(c) [bré.èd]	*!	*	
(d) [bú.rèd]		*!	*DEP

The tableau in (6) shows that there is no ranking between \*COMPLEX and NoCODA, since all output forms satisfy them. A candidate needs only to violate one of them for the form to be ruled out by the grammar. The Pre-Crowther ranking argument in (6) appears to be extremely active in the earlier stage of contact between English and Yoruba, as well-established nativized loanwords like *búrédì* in (6.III) show. We see in (6.III) that clusters and codas in English loanwords are resolved by epenthesis to satisfy the two high-ranking constraints. The ranking argument in (6) is thus motivated: even though the winning candidates violate FAITH, they emerge as the winner for obeying both top-ranking constraints. This ranking adjusts the structures of loanwords coming into Yoruba to the phonotactic rules of the language, ensuring that, in this specific case, markedness of the target is more important than faithfulness to the source. However, the constraints in the tableau in (6) have now been re-ranked to generate new optimal candidates, and formerly inactive constraints have been re-ranked to an active position so that some of the ungrammatical candidates in Pre-Crowther Yoruba are now parsed as grammatical. A discussion of these emerging phonological forms follows in the next section.

**3. EMERGENCE OF NEW GRAMMARS IN CONTEMPORARY YORUBA PHONOLOGY**

In this section, I show that diachronic changes in Yoruba phonology have already taken place, leading to synchronic variation in the realization of clusters and codas. In the previous section, I discussed both native and loan words in Yoruba that contain neither clusters nor codas. It has been suggested that the Pre-Crowther Yoruba phonological lexicon is characterized by the absence of codas and clusters.

I assume that this generalization is also true of the earlier stage of the English–Yoruba contact, judging from the form of the well-established English loans in the language, and the fact that researchers did not begin to notice sporadic cluster and coda realization before the early nineties (see Ufomata 1991 and Adeniyi 2015). If these assumptions are accurate, it follows that the Pre-Crowther Yoruba phonological lexicon is unstratified along the lines of clusters and codas. In Contemporary Yoruba, however, a diachronic stratification along the line of codas and clusters appears to have taken place. For ease of exposition I now present the evidence, assuming that special attention to input/output correspondence is necessary. Similarly, and for ease of discussion of the OT approaches to variation discussed in the next section, I refer to each input/output correspondence as Grammar 1, Grammar 2, etc. I provide a few examples for each pattern; more examples for each pattern can be found in Appendix B.

(7) The Core: Grammars 1–4

	Input	Output				Gloss
		Grammar 1	Grammar 2	Grammar 3	Grammar 4	
a.	brégèèd	búrégèèdì	brégèèdì	búrégèèd	brégèèd	be elaborate
b.	exam	èsåàmù	èksáámù	èsåám	èksáám	
c.	bread	búrèdì	brèdì	búrèd	brèd	
d.	slippers	sílípáàsì	slípààsì	sílípààs	slípààs	

In Grammar 1, clusters and codas are disallowed, and their resolution is done via epenthesis with a high vowel or by deletion. In Grammar 2, however, clusters are left unresolved while codas are resolved. Grammar 3 is a cluster-resolving but coda-retaining grammar. In Grammar 4, both clusters and codas are preserved. Grammars 1–4 exemplified in (7) above form the core of the Yoruba stratified lexicon in that this stratification is extremely productive in the grammar of all contemporary Yoruba monolingual and English–Yoruba speakers of Yoruba language, regardless of age, gender, or geographic region. Sociolinguistic stratification is thus a matter of quantitative variation, rather than a categorical one where a sociolinguistic category may preclude use of clusters or codas. Sociolinguistic variation is, however, outside the scope of this article, but has been addressed in Adebayo (2022).

In addition to these four grammars, there are three more, taken to form the Periphery in the Contemporary Yoruba phonological lexicon, as they only apply to a (growing) number of native and loan words. The unifying characteristic of the Periphery grammars exemplified in (8) is that they all involve hypercorrection of either consonant clusters or codas. That is, codas or clusters that are absent in the input are realized in the output. Hypercorrection is a well-documented phenomenon in variationist sociolinguistics since Labov (1966). While it can refer to people using a particular prestigious feature more than expected (called *quantitative hypercorrection* by Janda and Auger 1992), hypercorrection in this article is what these authors refer to as *qualitative* hypercorrection, where a prestigious linguistic feature not found in the input surfaces in the output (see similar examples in Winford 1978, Chung 2006, Chappell 2014). Grammars 5 and 6 are illustrated here (and in



Appendix B) with only English loans because no native Yoruba word has been found to simultaneously exhibit cluster hypercorrection and coda retention/coda hypercorrection. Grammar 5, as shown in (8), creates clusters not found in the input while retaining codas found in the input. That is, there is both cluster hypercorrection and coda retention. Also shown in (8), Grammar 6 (coda-resolving, cluster hypercorrection grammar) is so called because despite hyper-corrected clusters in the output, codas in the input are resolved.

(8) Cluster hypercorrection grammars: Grammars 5 and 6

	Input	Output	
		Grammar 5	Grammar 6
a.	president	présdéñt	présdéñtì
b.	jubilate	Jublèèt	Jublèètì
c.	formulate	fòmllèèt	fòmllèètì
d.	paracetamol	parastamòṛl prastamòṛl	parastamòṛ prastamòṛ

Note that in (8d), the word *paracetamol* is adapted as *parastamòṛ* with one instance of cluster hypercorrection and *prastamòṛ* with two instances of cluster hypercorrection. This shows that there can be two instances of cluster hypercorrection within a single word, an instance of what Kimpa (2011) calls local variation. In (8a), the word *president* is adapted as *présdéñt* and *présdéñtì*. Cluster hypercorrection (*sd*) takes place even though a cluster (*pr*) is already present in the input. This also suggests that cluster hypercorrection can take place in a form already containing a cluster. Grammar 7, exemplified in (9), consists of mainly native words where codas are hypercorrected. Loan words with coda hypercorrection are rare (but see Appendix B for a few examples).

(9) Coda hypercorrection grammar: Grammar 7

	Input	Output	Gloss
a.	Şàgámù	Şàgáàm	name of a city
b.	Oşòdì	Oşòòd	an area in Lagos
c.	sáàşì	sáàş	search
d.	ìgbádùn	ìgbaàd	enjoyment

We can now summarize the major generalizations that emerge from the data presented in (7) through (9) and in Appendix B. First, we have seen that Yoruba now not only permits clusters and codas but is creating clusters and codas where they did not originally exist. Since hypercorrection has been documented to be involved in language change (e.g., Janda and Auger 1992: 202), the hypercorrection data in examples (8 and 9) and in Appendix B provide further support for an important claim of this article: that Yoruba now permits clusters and codas. Hypercorrection examples are evidence that the phonotactic rules that Yoruba is borrowing from English through loanwords have been phonologized in the synchronic grammar of Yoruba speakers. The implication of this is that Yoruba speakers are not only borrowing segments from English (e.g., the

borrowing of /p/ (Ufomata 1991)) but are also borrowing phonotactic rules from the language.

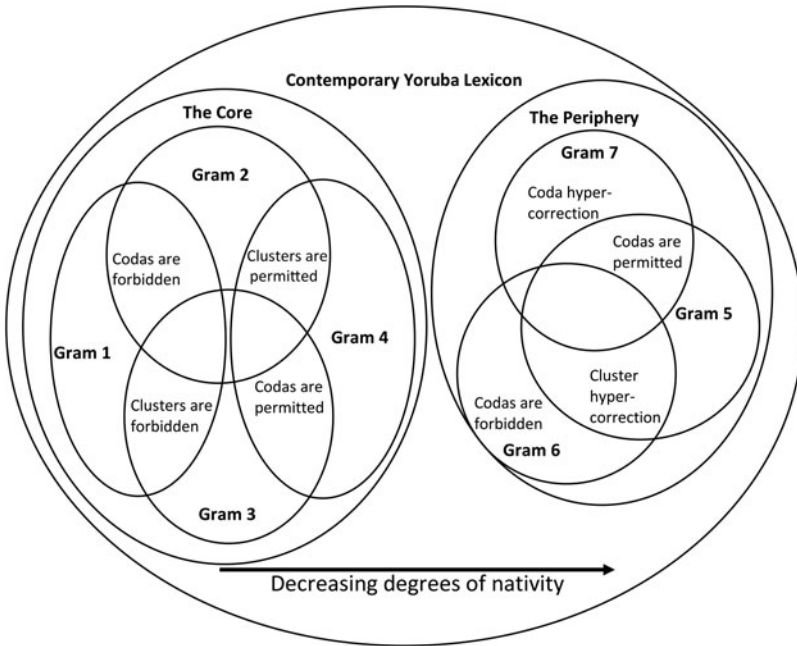
Second, an interesting generalization that emerges from the data described above is that cluster and coda hypercorrections are achieved mainly by syncope or apocope of high vowels /i/ and /u/.<sup>4</sup> Interestingly, these are also the two vowels that Yoruba uses to resolve clusters and codas. This suggests that hypercorrection follows a well-organized pattern: non-high vowels are more harmonic than high vowels, and so high vowels are deleted at a much higher rate than non-high vowels. This is approximately the harmonic scale proposed for Yoruba vowels in Pulleyblank (2008): Non High > High Back > High Front. Going by Steriade's (2009) P-map hypothesis, it can be suggested that high vowels are closer to zero than non-high vowels. This would explain why high vowels are the default epenthetic vowels (see Pulleyblank's 2008 redundancy rules), and why they are deleted more in hypercorrection and other processes such as vowel hiatus resolution.

There also appears to be a harmonic hierarchy when it comes to hypercorrection, as cluster hypercorrection and coda hypercorrection do not operate at the same rate. Also, processes such as vowel deletion and consonant epenthesis which lead to hypercorrection do not operate at the same pace. Cluster hypercorrection appears to take place at a much higher rate than coda hypercorrection as can be seen in Appendix B, where there are more examples of cluster hypercorrection than coda hypercorrection. Syncope is, thus, more likely than apocope. This means that vowel deletion in word-final position is more costly than in less perceptually salient word-medial position. Interestingly, /s/ is primarily involved in the consonant insertion hypercorrection forms. What is even more interesting is that hypercorrection takes place word-finally at a much higher rate than word medially: only one instance of word-medial consonant insertion is attested in Appendix B (brèskóyò), whereas there are several word-final examples, in direct contrast to vowel deletion.

Third, a major distinction can be made between the Core of the Yoruba phonological lexicon and its Periphery. The Core is characterized by free variation. An input with a cluster and a coda comes out as four optimal forms: one in which the cluster and the coda are resolved (Grammar 1), another in which the coda is resolved but the cluster is retained (Grammar 2), another in which the cluster is resolved but the coda is retained (Grammar 3), and a final one in which both the cluster and the coda are retained (Grammar 4). The Periphery is characterized by hypercorrection of clusters (Grammars 5 and 6) and codas (Grammar 7) where such hypercorrection does not exist for the Core. The distinction between the Core and the Periphery can further be illustrated as in the following diagram:

---

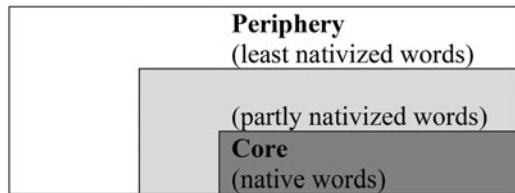
<sup>4</sup>Examples of hypercorrection not involving high vowels include *gbèra* 'immediately', *motèsaiku* 'motorcycle', *kérésimesi* 'Christmas', *èkáábè* 'welcome', *kpor* 'be plenty' (consonant /r/ is paragogized), and *wobiis* 'look here', *oyaas* 'it is time', *iyákókós* 'woman who sell kókó', and *wáás* 'come' (consonant /s/ is paragogized).



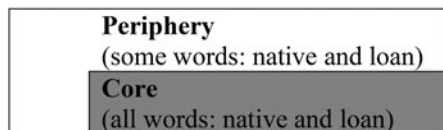
**Figure 1:** Contemporary Yoruba Phonological Lexicon

The illustration in [Figure 1](#) suggests that, unlike the Pre-Crowther Yoruba, where categorical choices are made against clusters and codas, the contemporary Yoruba phonological lexicon is one characterized by degrees of permissibility. This pattern, however, differs from the core–periphery structures reported so far for natural language. Hsu and Jesney (2018) report that there are three patterns found in the core–periphery model of the lexicon displayed in (10a).

(10) a. The core–periphery model of the lexicon (Hsu and Jesney 2018)



b. Contemporary Yoruba core–periphery lexicon



A major characteristic of the core–periphery model of the lexicon is that phonological generalizations that hold in the core ‘vary in how far they extend into the less nativized periphery’ (Hsu and Jesney 2018, after Holden 1976). It is generally assumed for the core–periphery model in (10a) that native words occupy the core where markedness restrictions native to a language are strictly enforced. Non-native words, on the other hand, occupy the periphery, where the restrictions become weaker and weaker as one moves from one stratum to another. Hsu and Jesney (2018) have shown that, in addition to this ‘superset at periphery’ pattern, there are at least two other patterns found in natural language: one in which markedness restrictions that are enforced on loanwords are inactive for native words (‘subset at periphery’) and another one in which marked structures are resolved differently in the core and the periphery (‘divergent repair’). What we have in the Yoruba data captured by (10b), however, differs from the native-in-the-core/foreign-in-the periphery pattern of (10a) in that phonological generalizations that hold in the core hold of all words whether they are a native word like *brégèèd* ‘be elaborate’ (→ *brégèèd*, *brégèèdì*, *bùrègèèd*, *bùrègèèdì*) or a loanword like *bread* (→ *brèèd*, *brèdì*, *bùrèèd*, *bùrèdì*), while phonological generalizations that hold in the periphery apply to some words whether they are native like *gbègìrì* (→ *gbegri* ‘a kind of soup’) or a loan word like *supermarket* (→ *spamakeṛet*). This suggests that the pattern documented in Hsu and Jesney (2018) in (10a) is not the only way by which a core–periphery structure can be organized. In section 4, I provide an Optimality-theoretic account of the pattern in (10b), reviewing some current OT approaches to diachronic changes, lexical stratification, and synchronic variation. I show in section 4.5 that adopting a lexically indexed MaxEnt model enables us to properly identify the stratification depicted in (10b) as a special kind of the ‘superset at periphery’ pattern.

Fourth, I have suggested that the changes reported in this section are due to influences of English. But the question could be asked whether clusters and codas existed already in some ideophone stratum of Pre-Crowther Yoruba phonology. As far as I am aware, no work has documented clusters and codas for Yoruba ideophones. Though Ola Orié (p.c.) suggests that oral rendition of the Bata drum may have clusters like *krákrá*, there is evidence to suggest that presence of clusters and codas in Yoruba ideophones is an invention of Contemporary Yoruba: (1) some Pre-Crowther Yoruba ideophones now undergo cluster hypercorrection: *fúkèfúkè* → *fkéfké* ‘rising spontaneously’, *gbrigidì* → *gbrigidi* ‘a manner of rolling on the floor’, *sírísírí* → *srísrí* ‘flowing in bits’; (2) ideophones with clusters and codas are recent neologisms: *gbaas-gbòós* ‘fracas’, *flapàpà* ‘fumble’. In sum, there is as yet no evidence, as far as I am aware, to suggest that clusters and codas existed in Pre-Crowther Yoruba or that a Pre-Crowther ideophone stratum is responsible for the variation that now characterizes Contemporary Yoruba syllable structure.

#### 4. OPTIMALITY-THEORETIC (OT) ANALYSES

Before we provide an OT account, we need to re-assess the set of constraints that we are working with. In section 2, we see that three major constraints are sufficient to capture the syllable structure of Pre-Crowther Yoruba phonology. We made use of

NoCODA, \*COMPLEX, and of course FAITH, which is an abbreviation for [MAX and DEP]. Given the data presented in the previous section, it is necessary to update this constraint set if we are to provide a useful account of the Contemporary Yoruba phonology. NoCODA, \*COMPLEX, and FAITH continue to be relevant, but there are additional constraints that are motivated by the data in section 3, which must be added to the set.

In Appendix B, it can be seen that the loanword ‘Sprite’ is adapted as *spráit*, *spráiti*, *sípráit*, and *sípráiti*. The forms *sípráiti* and *sípráit* stand out in that they suggest that a constraint against a cluster of more than two consonants is active in the grammar. There are other examples like *sròḡ* ‘straw’, *sréèt* ‘straight’, etc. for which this constraint seems to be active. This constraint is formalized as in (11a). Coda and cluster hypercorrections in examples (8–9) also suggest that we need markedness constraints that penalize candidates without clusters and those without word-final codas. As a heuristic, I formalize the constraint that penalizes candidates without clusters as in (11b) and follow McCarthy (2008) in assuming the existence of a constraint that penalizes candidates without word-final codas as shown in (11c).

- (11) a. \*CCC: triconsonantal sequences are forbidden (see, e.g., Pater 2007).  
 b. CC-SEQUENCE (CC-SEQ): a word without at least a sequence of two consonants (tautosyllabic or hetero-syllabic) is forbidden.  
 c. \*V#: ‘assign one violation mark for every word that ends in a vowel’ (McCarthy 2008: 171)

I refer to the constraints in (11b–c) as hypercorrection markedness constraints to highlight the fact that they are direct opposites of \*COMPLEX and NoCODA, respectively. Constraints (11b–c) can be motivated on two independent grounds. First, this is not the first time that a constraint that directly opposes well-established constraints has been proposed, and there is typological evidence for both constraints. \*V# is used by McCarthy (2008: 8–12) to explain data from Yawelmani. Itô and Mester (1995a) also propose two opposing constraints, PAL and DEPAL, to account for palatalization in the context of front vowels. Although they later offer an analysis in Itô and Mester (2003) that replaces DEPAL with CONTRAST, allowing directly opposing markedness constraint will depend on one’s ultimate theory of CON. There is nothing intrinsic to OT against such constraints. Typological evidence can also be invoked in support of \*V#. Blevins (2006: 336) reports that in languages like Olgol/Oykangand, Dvaravati Old Mon and language branches like the modern Palaungic, Khmuic, and Aslian, all words end in a consonant. This suggests that the postulation of \*V# is motivated on typological grounds. This means that even though it is common for languages (e.g., Pre-Crowther Yoruba) to forbid word-final codas, there are languages that prefer the opposite. Also, although it is uncommon for languages to forbid structures without clusters, there is documented evidence that some languages create clusters in output forms that are not present underlyingly, not to satisfy any particular well-formedness condition but due to social pressures. Specifically, Chappell (2014: 33, 38) reports that Dominican Spanish speakers insert pre-cluster and pre-coda /s/ as a marker of prestige. This means that while it is almost impossible for a language to independently require the presence of consonant clusters in optimal forms, social pressure could

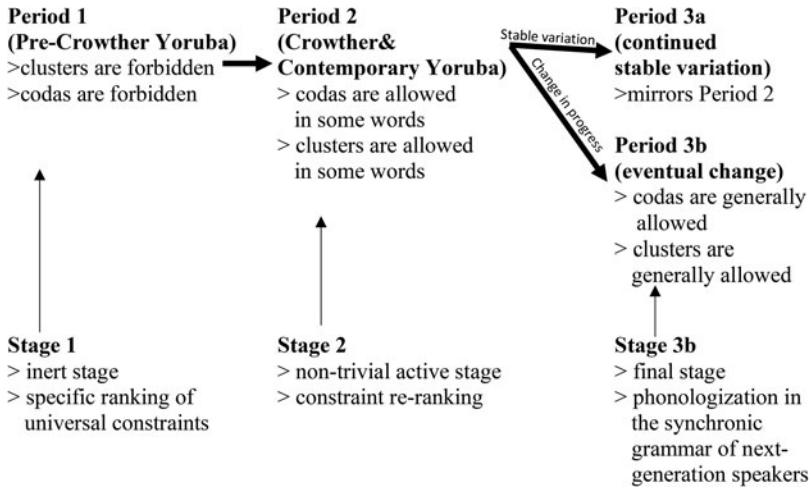
lead to just this situation. CC-SEQ can then be said to be motivated by the consequences of social pressure on linguistic forms.

Second, there are no well-established constraints that fill the role of \*V# and CC-SEQ in the analyses that follow. Since most of the hypercorrection cases outlined in section 3 involve deletion of high vowels, one may be tempted to propose a constraint like \*HI against high vowels; but we have seen that both low and mid vowels also are deleted, in some words. Even if we posit \*LOW and \*MID, we cannot account for cases where hypercorrection arises as a result of consonant insertion (e.g., brèkóyòn → brèskóyòn ‘bra’; wòó → wòòs ‘look here’). A reviewer suggests that perhaps MAX-V can account for cluster and coda hypercorrection, since hypercorrection involving vowel deletion is more frequent than those involving consonant insertion. Such an analysis, however, still does not account for hypercorrections involving consonant insertion. For this reason, I do not pursue this analysis. Instead, I argue that \*V# and CC-SEQ are sufficiently motivated. With the relevant set of constraints defined, I turn now to providing an OT account.

#### 4.1 Diachronic changes in Yoruba as constraint re-ranking

One of the fundamental principles of OT is that Universal Grammar (UG) consists largely of well-formedness constraints out of which individual grammars are constructed (Prince and Smolensky 2004 [1993]. This means that different rankings of the same universal constraints yield different grammars of different languages (McCarthy 2008, etc.). For instance, English and Yoruba have the same constraints such as NOCODA, FAITH, and \*COMPLEX. The difference between the two languages is simply in how these constraints are ranked. This principle successfully accounts for cross-linguistic variation among languages. However, one implication of this theoretical assumption is that different speakers of the same language may employ different rankings of the same constraints across social contexts and points in time. OT, therefore, should be able to account for both speaker variation and diachronic variation. This, in fact, is the position of such works as Jacobs (1995 and 1996), Gess (1996), Hutton (1996), Holt (1997), Miglio and Moren (2003). Miglio and Moren (2003: 192), in particular, observe that “since grammars are conceived of as specific ranking of the same universal constraints, it follows that language change must be a re-ranking of those universal constraints.” Although language change is not a formulated part of original OT, its theoretical validity follows straightforwardly from its main assumptions.

To account for changes in the syllable structure of Yoruba phonology, I follow Miglio and Moren’s (2003) framework for language change within OT. For them there are three stages of language change: the inert stage, the second stage, and the final stage. At the inert stage, the traditional or language-specific ranking of the universal constraints is maintained. At the second stage, at least one of the constraints has been re-ranked, while at the final stage, constraint re-ranking has been phonologized in the synchronic grammar of the next generation of speakers. The diagram in Figure 2, incorporating Miglio and Moren’s (2003) three stages of change in OT, gives a visual description of the timeline of the changes taking place in Yoruba and in possible future trajectories.



**Figure 2:** Diachronic changes in Yoruba syllable phonology

Stage 1 corresponds to the Pre-Crowther Yoruba phonology as described in section 2. I argue that the period of time between Crowther Yoruba and Contemporary Yoruba corresponds to the second stage, where at least one of the active constraints in Yoruba syllabic phonology has been re-ranked, leading to variation in cluster and coda realization. The third stage of the change, where syncope and apocope of vowels (leading to coda and cluster hypercorrection) occur without lexical stratification, is not yet attested. For this reason, I focus on the second stage, by providing a synchronic account of the variation in line with current OT approaches to variation; a variationist study of this sociolinguistic situation can be found in Adebayo (2022).

Before proceeding to an OT account of the variation described in section 3, let us sum up. Stage 1, Pre-Crowther Yoruba, is an inert stage characterized by the following ranking argument: \*COMPLEX, NoCODA >> FAITH. The relevant constraint set is {NoCODA, \*COMPLEX, FAITH}. Following Miglio and Moren (2003), re-ranking these constraints should describe all the data outlined in section 2. However, no ranking argument containing only these three constraints suffices to account for the variation discussed in section 3. In addition to re-ranking, the set of active constraints whose effects manifest in a synchronic grammar has expanded. This, therefore, means that some diachronic changes may not result solely from constraint re-ranking, but also from the expansion of the set of active constraints. The idea is that a language has two sets of constraints at any given point: a) a set of active constraints which shape phonological outcomes and b) a set of inactive constraints whose effect cannot be seen in the language at a given point in time. Expanding the set of active constraints in a synchronic grammar entails re-ranking since formerly inactive constraints are re-ranked to be active, but re-ranking does not entail expansion, since the same set of active constraints

can be re-ranked to yield new outputs. While \*COMPLEX, NoCODA >> FAITH has been re-ranked, the set of active constraints has also been expanded as a result of the fact that three formerly inactive constraints, \*CCC, CC-SEQ, and \*V#, have been re-ranked so that they are now a member of the set of active constraints.

There are four main OT approaches to synchronic variation and lexical stratification considered in this article. These are: the ranked-winners approach (Coetzee 2004), the indexed constraint approach (Itô and Mester 1995a, b and 1999; Fukazawa 1998), the partial-order multiple grammars approach (Anttila 1997, 2002; Anttila and Cho 1998), and the MaxEnt model of Goldwater and Johnson (2003). I review each of these four approaches against the Yoruba data in section 3 and show that none of them account for it without invoking further mechanisms. I demonstrate in section 4.5, however, that the most economical model, requiring minimal ad hoc stipulations, is the MaxEnt model.

## 4.2 Yoruba data and the ranked-winners approach

A major assertion of Coetzee's (2004) is that the same ranking argument can generate more than one optimal candidate. Most OT tableaux in the literature have only one winning candidate – even when two winners are allowed, they must have equal number of violations (McCarthy 2008). However, we know from language variation and change research that there is often more than one way of saying the same thing, so that, for example, an OT account of *in/ing* in some varieties of English (e.g., Mendoza-Denton 1997) must allow for the emergence of more than one optimal candidate. This follows straightforwardly if we pursue OT's assumptions to its logical conclusions. The assumption that the winning candidate emerges by incurring fewer violations on high-ranking constraints than losing ones have done implies that the candidates generated by GEN can be arranged on a hierarchy by EVAL so that the winning candidate is the most harmonic, while the candidate that incurs second-fewest violations is the second most harmonic, and so on. This is the conclusion reached in Coetzee (2004), who proposes that EVAL imposes 'a harmonic rank-ordering' on candidate sets so that some losers are better than others. In this model, there is room for more than one optimal candidate, which can be arranged on a scale such as the following: most optimal >> second most optimal >> and so on. The implication of this is that losers have the potential to become optimal in a synchronic grammar, but the grammar specifies the 'critical cut-off point' on the constraint hierarchy (Coetzee 2004: 366), which determines the range of optimal candidates in a candidate set.

Coetzee's approach, however, runs into some serious problems with the Yoruba data by the simple fact that a single ranking argument cannot generate all the optimal outputs described in section 3. As we will see below, some form of constraint re-ranking is needed. Consider the following tableau (for purposes of illustration and simplicity, I omit \*CCC from the discussion):



(12) One ranking argument, multiple winners

		MAX	NoCODA	*COMPLEX	DEP	CC-SEQ	*V#
I. /brɛd/ <sup>5</sup>	☞ (a) [bú.rɛ̀.ði]				**	*	*
	☞ (b) [brɛ̀.ði]			*	*		*
	☞ (c) [brɛ̀.ɛ̀d]		*	*			
	☞ (d) [bú.rɛ̀.ɛ̀d]		*		*	*	
	(e) [bɛ̀d]	*!	*			*	
	(f) [bɛ̀ði]	*!			*	*	*
II. /dʒu:biɛɪt/	☞ (a) [ju.bi.lé.èt]		*			*	
	☞ (b) [ju.bi.lé.é.ti]				*	*	*
	☞ (c) [ju.blé.èt]	*!	*	*			
	☞ (d) [ju.blé.é.ti]	*!		*	*		*
	(e) [ju.bi.lé]	*!				*	*
	(f) [ju.blé]	*!*					*
III. /bɛ̀ɛ̀ ni/	☞ (a) [bɛ̀.ɛ̀.ni]					*	*
	☞ (b) [bɛ̀.ɛ̀n]	*!	*			*	
	(c) [bɛ̀.ɛ̀]	**!				*	*

Let us start from this observation: with a given input like /brɛd/ (12.I), the optimal candidates can be arranged on a scale that depicts decreasing degrees of nativity (i.e., native ([bú.rɛ̀.ði]) → less native ([brɛ̀.ði]) / ([bú.rɛ̀. ɛ̀d]) → least native ([brɛ̀.ɛ̀d])). Employing Coetzee’s ranked-winners grammar entails that we must find a ranking argument that ensures precisely this harmonic rank-ordering. Insofar as we have something close to this harmonic rank-ordering, the ranking argument in (12) is motivated, to some heuristic extent.

According to Coetzee (2004: 23), language users do not access a candidate that is disfavored by a constraint above the critical cut-off point, if there is a candidate that is

<sup>5</sup>While I assume that the lexical representation of loans are in their original forms, there is evidence for their restructuring. Evidence can be seen in cluster hypercorrection of forms such as ‘àbùradà’, restructured from the lexical representation for ‘umbrella’. We know this because ‘àbùradà’ is the input for the hypercorrected form ‘àbradà’. However, when the lexical representation of a loan is restructured, both the original representation and the restructured form become part of the lexicon, and ‘umbrella’ remains the input for ‘òmbùrɛ̀là’, ‘òmbùrɛ̀là’, ‘òbrɛ̀là’, ‘òbùrɛ̀là’.

not disfavored by such a constraint. In the tableau in (12.I), the critical cut-off point (indicated by the double vertical lines) is drawn in the MAX column, indicating that candidates [béd] and [bè.dì] (violating MAX) are not accessed by the language user. As expected, the grammar generates four optimal outputs (a, b, c, and d) in (12.I). As it stands, [bú.ré.dì] is the optimal candidate while [bré.dì] is the second best candidate. [bú.ré.èd] is the third best candidate, while [bré.èd] is the fourth best candidate. This harmonic rank-ordering contrast with the expected rank ordering ([bú.ré.dì] (optimal) → [bré.dì]/ ([bú.ré.d] (second best) → [bré.èd] (third best)). This problem is trivial, given that we can ensure the correct rank-ordering by assuming that there is no ranking between \*COMPLEX and NoCODA. The cluster hypercorrection example in (12.II) reveals a different kind of problem. Even though the grammar produces the correct rank-ordering ([ju.bi.lé.é.tì] → [ju.bi.lé.èt] → [ju.blé.é.tì] → [ju.blé.èt]), not all the winning candidates are correctly predicted to be optimal.

The coda-hypercorrection example in (12.III) too has the correct harmonic rank-ordering ([bè.è.ni] → [bè.èn]). The problem with this example is that the coda-hypercorrection optimal candidate ([bè.èn]) is predicted to be impossible since it violates MAX, a constraint above the critical cut-off point. Apart from all the various problems identified so far, the ranked-winners approach also does not provide us with a straightforward way to capture the core-periphery organization inherent to the Contemporary Yoruba phonological lexicon depicted in (10b).

### 4.3 Yoruba data and the indexed constraint approach

Let us consider next the indexed constraint approach to synchronic variation/phonological stratification (Itô and Mester 1995a, b, 1999; Fukazawa 1998; Pater 2000). According to Pater (2004), there are different versions of this approach: those in which lexical items are indexed with specific ranking of two or more constraints (McCarthy and Prince 1993, Nouveau 1994, Pater 1994, etc.), and those in which constraints are indexed to some particular lexical stratum or morphological category (Itô and Mester 1995a, b, 1999; Fukazawa 1998; Pater 2000). These two approaches are further divided into different variants, but I focus only on the latter approach, specifically that advanced in Itô and Mester (1995a, b, 1999).

Itô and Mester (1995a, b, 1999) observe that the phonological lexicon of Japanese is stratified along the lines of lexical origin so that different phonological processes characterize different groups of words. They posit four strata for Japanese: Yamato, Sino-Japanese, Assimilated Foreign and Unassimilated Foreign. This stratification is organized in terms of a core-periphery structure in the following fashion: Yamato → Sino-Japanese → Assimilated Foreign → Unassimilated Foreign. Movement from Yamato to Unassimilated Foreign correlates with movement from the core to the periphery, similar to what I have described in section 4.2 as ‘decreasing degrees of nativity’. In order to retain the ‘crucial ranking’ assumption of classical OT, Itô and Mester (1999) propose that Japanese can be accounted for with a fixed

ranking of markedness constraints, with the re-ranking of only faithfulness constraints at different lexical strata. As we move from the core to the periphery, more and more markedness constraints are violated by optimal candidates in peripheral strata. Let us illustrate this with the Yoruba data in the following tableau.<sup>6</sup> Shaded columns indicate that a given re-ranking of FAITH is inactive. In this illustration, Stratum 1 corresponds to Grammar 1, Stratum 2 to Grammar 2, etc.

(13) Indexed constraint tableau

		FAITH <sub>S4</sub>	NoCODA	FAITH <sub>S2</sub>	*COMPLEX	FAITH <sub>S1</sub>	CC-SEQ	FAITH <sub>S4S3</sub>	*V#	FAITH <sub>S6</sub>
I. /bred/s <sub>4</sub>	(a) [bú.rẹ̀.dì]	*MAX!					*		*	
	(b) [brẹ̀.dì]	*DEP!			*				*	
	(c) [brẹ̀.ẹ̀d]		*		*					
II. /bred/s <sub>2</sub>	(a) [bú.rẹ̀.dì]			**DEP!			*		*	
	(b) [brẹ̀.dì]			*DEP	*				*	
	(c) [brẹ̀.ẹ̀d]		*!		*					
III. /bred/s <sub>1</sub>	(a) [bú.rẹ̀.dì]					**DEP	*		*	
	(b) [brẹ̀.dì]				*!	*DEP			*	
	(c) [brẹ̀.ẹ̀d]		*!		*					
IV. /dzu:biłet/s <sub>5</sub>	(a) [ju.bi.lé.èt]		*!				*			
	(b) [ju.bi.lé.è.ti]						*	*DEP	*	
	(c) [ju.blé.èt]		*!		*			*MAX		
	(d) [ju.blé.è.ti]				*!			*MAX, *DEP	*	
V. /dzu:biłet/s <sub>6</sub>	(a) [ju.bi.lé.èt]		*!				*			
	(b) [ju.bi.lé.è.ti]						*	*DEP	*	
	(c) [ju.blé.èt]		*!		*			*MAX		
	(d) [ju.blé.è.ti]				*!			*MAX, *DEP	*	
VI. /bẹ̀ɛ̀ni/s <sub>7</sub>	(a) [bẹ̀.ẹ̀.ni]						*		*	
	(b) [bẹ̀.ẹ̀n]		*!				*			*MAX
	(c) [bẹ̀.ẹ̀]						*		*	**MAX!

I first assume, based on the lexical organization in Figure 1, that contemporary Yoruba phonology is characterized by the crucial ranking of markedness constraints in (14a). I then further assume that FAITH is re-ranked at each stratum as in (14b).

<sup>6</sup>I leave out \*CCC here to focus on the issue at hand, as well as Grammar 3 since including it will create problems from the start, thereby preventing the illustration. I return to this problem in the next two paragraphs.

(14) a. NoCODA >> \*COMPLEX >> CC-SEQ >> \*V#

b. FAITH<sub>S4</sub> >> NoCODA >> FAITH<sub>S2</sub> >> \*COMPLEX >> FAITH<sub>S1</sub> >> CC-SEQ >> FAITH<sub>S5/6</sub> >> \*V# >> FAITH<sub>S7</sub>

Each re-ranking of FAITH corresponds to a stratum. The stratification is thus FAITH-based (Inkelas and Zoll 2007:135). This approach works perfectly for the Core of Yoruba phonological lexicon (if Stratum 3 is not included). Indeed, as we move from Stratum 1 to Stratum 4 in (13), more and more markedness constraints are violated by optimal candidates in the strata. In (13.III), exemplifying Grammar 1, the coda-resolving, cluster-resolving candidate ([bú.r̥.ɗì]), which obeys both of the markedness constraints (NoCODA and \*COMPLEX) ranked above FAITH<sub>S1</sub>, wins against both the coda-retaining and cluster-preserving candidate ([br̥.ɛ̀.ɛ̀]) and the coda-resolving, cluster-preserving candidate ([br̥.ɗì]) which violate either of the two constraints. As we move from Stratum 1 to Stratum 2 (13.II), we encounter a grammar in which the optimal candidate [br̥.ɗì] violates one of the markedness constraints (\*COMPLEX) which the optimal candidate in Stratum 1 must not violate. In Stratum 4, we find a grammar in which the optimal candidate [br̥.ɛ̀.ɛ̀] violates both NoCODA and \*COMPLEX. The Core of Yoruba phonology (minus Stratum 3), based on this characterization, is thus a perfect example of what Itô and Mester (1999: 62) describe as nesting of constraint domains.

The trouble with this approach, however, begins when we incorporate Stratum 3 in the analysis. As the FAITH-based re-ranking in (14b) stands, there is no way to generate Stratum 3 optimal forms without re-ranking \*COMPLEX so that it outranks NoCODA. Since re-ranking of markedness constraint is not allowed in this approach, we have no way to account for Stratum 3. We encounter further problems as we move to the Periphery of the Yoruba lexicon. First, the re-ranking of FAITH is problematic. There is no way to rank it with respect to the four markedness constraints such that it yields all seven strata. The problem is that Grammars 5 and 6 are distinguished only with respect to whether or not coda is resolved in a cluster-hypercorrection form. Since NoCODA is already present, choosing between these categorical choices, no new markedness constraint can be introduced such that it is active in either Stratum 5 or 6, and FAITH is re-ranked with respect to it. It also does not matter whether we include \*CCC or not. This approach, therefore, forces us to incorrectly conflate FAITH<sub>S5</sub> and FAITH<sub>S6</sub> into FAITH<sub>S5/6</sub> in (13). This conflation may not be damaging to the theory, if it were to correctly predict the optimal candidates in Grammars 5–7 as it does for Grammars 1–4 (excluding Grammar 3). This is not the case, however, as we see in (13.IV–VI). In each of Grammars 5–7 (13.IV–VI respectively), wrong candidates are predicted as winners, while optimal candidates are predicted to be losers. In (13.IV), exemplifying Grammar 5, the coda-resolving non-cluster-hypercorrection candidate [jubiléèti] incorrectly wins against the optimal coda-preserving cluster-hypercorrection candidate [jublèèt]. This scenario is exactly what we find in (13.V), where the same coda-resolving non-cluster-hypercorrection candidate [jubiléèti] wins against the optimal coda-resolving

cluster-hypercorrection candidate [jublèè̀t̩]. The same thing happens in (13.VI), where Candidate (a) incorrectly wins, and the optimal Candidate (b) loses.

That the FAITH-based indexed constraint approach runs into this conceptual problem with the Yoruba data is not surprising; we have already seen that any approach that does not allow re-ranking of markedness constraints is bound to encounter problems with the Yoruba data. The main difference between the Core of Yoruba lexical phonology and its Periphery is markedness reversal. The Core requires \*COMPLEX and NOCODA to be more active than the hypercorrection constraints CC-SEQ and \*V# at some point, while the Periphery requires the opposite. This fact of markedness reversal in the Yoruba data is thus a major problem for the FAITH-based indexed constraint approach of Itô and Mester (1995a, b, 1999). This problem of markedness reversal was in fact already identified by Inkelas and Zoll (2007:157–159). Despite these problems, Itô and Mester's (1995a, b, 1999) FAITH-based indexed-constraint approach offers useful insights on stratificational patterns found in the phonological lexica of natural languages. The idea that the lexicon of some languages is organized in a core–periphery structure allows us to construct an adequate theoretical explanation for the Yoruba data in section 4.5.

It should be clear by now that to account for the Yoruba data, we need an approach that allows some form of markedness constraint re-ranking which is constrained enough not to over-predict. As we have seen, the re-ranking of markedness constraints is important to capture the markedness reversal described above. In what follows, I show how the partial-order co-phonology model (Anttila 1997, 2002 and Anttila and Cho 1998) allows us to account for the markedness reversal found in the Yoruba data, although it ultimately runs into a major problem of its own.

#### 4.4 Yoruba data and the partial-order co-phonology

The partial-order co-phonology is a restrictive version of the general co-phonology/multiple grammars approach advanced in Kroch (1989), Kiparsky (1993), Orgun (1996), and Inkelas (1998). The general co-phonology approach assumes that a grammar is comprised of other sub-grammars/co-phonologies which are different as a result of different crucial rankings of the same optimality-theoretic constraints. Since these rankings are arbitrary (Anttila and Cho, 1998: 40), the general multiple grammars approach has the potential to generate unattested sub-grammars for a given language. The partial-order model of the multiple grammars approach proposes to solve this problem by assuming that constraints are partially ranked with respect to one another. In standard Optimality Theory, grammars are assumed to have the properties of irreflexivity (i.e., a constraint  $x$  cannot be ranked above or below itself), asymmetry (i.e., a constraint  $x$  which is ranked above another constraint  $y$  cannot be ranked below  $y$ ), transitivity (i.e., if a constraint  $x$  ranks above another constraint  $y$ , and  $y$  ranks above  $z$ , then the grammar also requires a ranking of  $x$  above  $z$ ) and connectedness (i.e., every constraint is crucially ranked with respect to every other constraint). The partial-order model assumes that the grammar of a language is

characterized by all of these properties except the property of connectedness (Anttila and Cho 1998: 36, Anttila 2002: 20). The main grammar (master constraint ranking) of an individual is thus partially determined so that the remaining rankings among all the active constraints are determined by the co-phonologies/sub-grammars. For convenience, let us follow Anttila (2007) and refer to the main grammar of an individual as GRAMMAR and the co-phonologies/ subgrammars as ‘grammars’. The GRAMMAR of an individual is not a random collection of grammars but a proper superset of all the grammars. The grammars are sets of ordered constraint pairs grouped into natural classes based on their shared rankings (Anttila 2009). Grammars are thus connected to one another just as they are connected to a GRAMMAR. This assumption enables the partial-order model to be able to capture the core–periphery structure that is one of the major hallmarks of the indexed-constraint approach described above. This relationship between a GRAMMAR and its grammars can be represented in a grammar lattice where higher nodes represent variant grammars and lowest nodes represent invariant grammars.

The Yoruba data in section 3 is better analyzed in the partial-order model than in the previous approaches discussed, but as we will see below, it runs into a major problem in that it requires too many ad hoc stipulations. First, I assume that the partial ranking that characterizes the Yoruba GRAMMAR is \*CCC>>\*COMPLEX. As we will see in the next section, CCC and CCCC clusters are consistently reduced to CC in appropriate contexts, suggesting that this ranking is motivated. Besides, ranking \*COMPLEX (a more general constraint) above \*CCC (its more specific counterpart) will constitute an anti-Paninian ranking for which we have no evidence in the Yoruba data. Assuming that this partial ranking defines the Contemporary Yoruba phonology, the grammar lattice in example (30) of Appendix C is assumed to represent the Yoruba GRAMMAR and its different grammars identified in section 3.

This grammar lattice presents a Yoruba GRAMMAR which contains different grammars that are all connected back to one main grammar in a core–periphery fashion. The lattice thus demonstrates that the dataset that we have seen in the previous section is not a collection of random facts about the Contemporary Yoruba phonology but, as Itô and Mester (1999: 69) observe, the result of a simple generalization that holds of any given stratified lexicon. One thing to note in each of the lowest leaves in the lattice is that they are not in themselves total rankings of all the six constraints that we are working with, such that every constraint is crucially ranked with respect to every other constraint, but rather partial rankings where the ranking between some constraints is not fixed. This means that the partial ranking in each of the lowest leaves is not just one tableau but a collection of several tableaux where the winner in each stratum will always emerge no matter how the remaining partial rankings are fixed. The major advantage of having several tableaux producing the winner in each stratum is that it introduces a quantitative paradigm into the system, which can be exploited to calculate the frequency of variants in the strata. Anttila and Cho (1998: 39) formalize this quantitative dimension as follows:

- (15) Quantitative interpretation of partially ordered grammar (Anttila and Cho 1998: 39)
- a. A candidate is predicted by the grammar iff it wins in some tableau.
  - b. If a candidate wins in  $n$  tableaux and  $t$  is the total number of tableaux, then the candidate's probability of occurrence is  $n/t$ .

This quantitative interpretation is often checked against usage data (actual frequency) acquired by sociolinguists, and it has proven successful so far (see Anttila 1997, 2002; Anttila and Cho 1998; and Zamma 2005).<sup>7</sup>

Let us now look at the grammars and how the lexical items in Yoruba relate to them. Abstracting away from the lexical indexation employed in Itô and Mester (1999), I assume that every lexical item in the phonological lexicon of Contemporary Yoruba is indexed with a grammar in which a variant of it is optimal. Consider *jubilate* and *parade* to illustrate why lexical indexation is important. *Jubilate* comes out as *jubilèè̀tì* (Grammars 1 and 2), *jubilè̀t* (Grammar 3, 4 and 7), *jublè̀tì* (Grammar 6), and *jublè̀t* (Grammar 5), whereas *parade* comes out as *pàrè̀dì* (Grammars 1 and 2) and *pàrè̀d* (Grammars 3, 4 and 7). Why are *\*prè̀dì* and *\*prè̀d* excluded as optimal variant of *parade*? We have the answer if we assume that the lexical entry for *jubilate* contains an indexation for Grammars 1–7, but that the entry for *parade* contains an indexation for only Grammars 1–4 and 7. Similarly, while a transformation such as *oşòdì* ('an area in Lagos') → *oşòd* is legitimate to create a coda hypercorrection form *oşòd*, a process such as *àwòdì* ('hawk') → *\*àwòd* is illegitimate. As such, *\*àwòd* is ungrammatical. Also, a transformation such as *sùgbòn* ('but') → *sgbòn* is legitimate to create a cluster hypercorrection form *sgbòn*, while a transformation such as *àşìgbò* ('mishearing') → *\*àşgbò* is illegitimate, making *\*àşgbò* to be ungrammatical. Given that *oşòdì* and *àwòdì*, on the one hand, and *sùgbòn* and *àşìgbò*, on the other hand, have similar phonological contexts favorable to coda hypercorrection and cluster hypercorrection respectively, one would expect that both these forms should undergo hypercorrection, but that is not the case. This expectation would have been borne out if hypercorrection were a general phonological rule applying across all relevant lexical items in the Contemporary Yoruba lexicon. We have a straightforward account if we assume that forms like *àwòdì* and *àşìgbò* are not (or probably not yet) indexed for Grammars 5–7. We then make the following observation: every lexical item in Yoruba (foreign or native) is indexed for Grammars 1–4, meaning that a variant of them will always win in each of Grammars 1–4, even if the variant is a single optimal form. Indexation for Grammars 1–4, therefore, is redundant, while active indexation takes place at the periphery. This observation becomes clearer in the light of the informal definitions below:

<sup>7</sup>A sociolinguistic dataset (aiming to uncover the sociolinguistic conditioning of the variation in section 3) is being prepared by the author; the present article provides a descriptive and analytical basis for such usage data; I therefore omit a discussion of the quantitative interpretation of the partial rankings in the lowest leaves in (30) (Appendix C).

## (16) Informal definitions of the seven grammars

- a. **Grammar 1 (cluster-/coda-resolving grammar):** if you are an input with a consonant cluster and/or a coda, your consonant cluster and coda are resolved. Otherwise, you come out as you enter the grammar.
- b. **Grammar 2 (cluster-preserving, coda-resolving grammar):** if you are an input with a consonant cluster and/or a coda, your consonant cluster is preserved while your coda is resolved. Otherwise, you come out as you enter the grammar.
- c. **Grammar 3 (cluster-resolving, coda-preserving grammar):** if you are an input with a consonant cluster and/or a coda, your consonant cluster is resolved while your coda is preserved. Otherwise, you come out as you enter the grammar.
- d. **Grammar 4 (cluster-/coda-preserving grammar):** if you are an input with a consonant cluster and/or a coda, your consonant cluster and your coda are preserved. Otherwise, you come out as you enter the grammar.
- e. **Grammar 5 (coda-preserving cluster hypercorrection grammar):** if you are an input with a coda but without a consonant cluster, your coda is preserved but you come out with a cluster. If you already contain a cluster, your cluster is preserved.
- f. **Grammar 6 (coda-resolving cluster hypercorrection grammar):** if you are an input with a coda but without a consonant cluster, your coda is resolved, and you come out with a cluster. If you already contain a cluster, your cluster is preserved.
- g. **Grammar 7 (coda hypercorrection grammar):** if you are an input with or without a word-final coda, you come out with a word-final coda.

Assuming that these informal definitions are accurate, we can now make the following claims:

- (17)a. All lexical items in Yoruba (foreign or native) are indexed for Grammars 1–4, while some lexical items are also indexed for Grammars 5 and 6 or Grammar 7.
- b. All lexical items indexed for either Grammars 5 and 6 or Grammar 7 are also indexed for Grammars 1–4 but not all lexical items indexed for Grammars 1–4 are also indexed for either Grammars 5 and 6 or for Grammar 7.

With the observations in (16) and (17) in place, we are now ready to examine how winners emerge in each of the strata. The tableau that follows (and those in Appendix D) are not just one single tableau but a conflation of several tableaux where a given winner always emerges. To be able to represent all of them at once, I have deployed the conventions in (18).

## (18) Conflation tableau conventions

- a. If a constraint X in a constraint set with the ranking  $[V \gg W \gg X \gg Y \gg Z]$  is annotated as ‘X...’ as in  $[V \gg W \gg X \dots \gg Y \gg Z]$ , this means that X is unranked with respect to constraints Y and Z even though V outranks every constraint in the set; W outranks X, Y, and Z; and Y outranks Z.
- b. If a constraint X in a constraint set with the ranking  $[V \gg W \gg X \gg Y \gg Z]$  is annotated as ‘...X’ as in  $[V \gg W \gg \dots X \gg Y \gg Z]$ , this means that X is unranked with respect to constraints V and W even though V outranks W, Y, and Z; W outranks Y and Z; X outranks Y and Z; and Y outranks Z.



- c. If a constraint X in a constraint set with the ranking [V>>W>>X>>Y>>Z] is annotated as ‘X (... V)’ as in [V>>W>>X (... V) >>Y>>Z], this means that even though all the rankings hold, X is unranked with respect to V.
- d. As usual, vertical broken lines and a comma between two constraints indicate that the constraints are unranked with respect to each other.

The convention in (18a) states that in a ranking where A>>...B>>C, both A>>B>>C and B>> A>>C are true, while the convention in (18b) says that in a ranking where A>>B...>>C, both A>>B>>C and A>>C>>B are true.

Based on the informal definitions in (16), Grammar 1 is a grammar that resolves both clusters and codas. As we see in (19.I), if an input contains a cluster or/and a coda, the cluster and the coda are resolved, and this is how Candidate (19.I.a) emerges as the winner. If, as we see in (19.II), the input does not contain a cluster or a coda, the most faithful candidate (19.II.a) emerges as the winner.

(19) Grammar 1: NoCODA>> \*V#...>> \*CCC>> \*COMPLEX>> FAITH, CC-SEQ

		NoCODA	*V#...	*CCC	*COMPLEX	FAITH	CC-SEQ
i. /bred/s1-7	☞ (a) [bú.rẹ̀.di]		*			**DEP	*
	(b) [brẹ̀.di]		*		*!	*DEP	
	(c) [brẹ̀.ẹ̀d]	*!			*		
ii. /oníbàtà/s1-4	☞ (a) [o.ní.bà.tà]		*				*
	(b) [on.i.bà.tà]	*!	*				*
	(c) [o.nbà.tà]		*		*!	**MAX	
	(d) [o.ní.bàt]	*!				**MAX	*
	(e) [o.nbàt]		*	*!	**	***MAX	

Given space constraints, I have limited the discussion of how winners emerge in our partial-order analysis to Grammar 1, but the tableaux for the remaining six grammars can be found in Appendix D. I turn now to summarizing the facts about Yoruba GRAMMAR with the examples in the following lexical stratification table (Table 1).

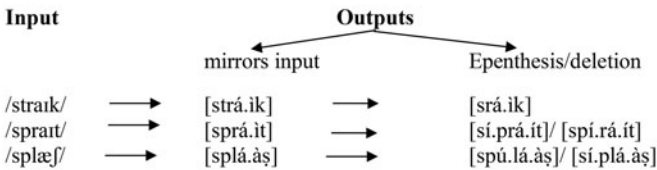
Some progress in capturing the Yoruba data within OT has been made by using a partial-order co-phonology, but this stands only because we have borrowed indexing from the indexed constraint approach. Despite this progress, the lexically-indexed partial-order analysis presented above cannot handle all the variation that currently characterizes Contemporary Yoruba phonology. Specifically, we are still unable to capture the variation in (20):

Input	Grammar 1	Grammar 2	Grammar 3	Grammar 4	Grammar 5	Grammar 6	Grammar 7
	Yoruba GRAMMAR						
	The Core				The Periphery		
/bred/s1-7	[bú.rẹ̀.di]	[brẹ̀.di]	[bú.rẹ̀.ẹ̀d]	[brẹ̀.ẹ̀d]	[brẹ̀.ẹ̀d]	[brẹ̀.di]	[brẹ̀.ẹ̀d]
/trei/ s1-6	[tí.rè]	[trè.è]	[tí.rè.è]	[trèè]	[trèè]	[trèè]	
/geit/ s1-4&7	[gé.è.ti]	[gé.è.ti]	[gé.èt]	[gé.èt]			[gé.èt]
/d̥ʒu:bileit/ s1-7	[ju.bi.lé.è.ti]	[ju.bi.lé.è.ti]	[ju.bi.lé.èt]	[ju.bi.lé.èt]	[ju.blé.èt]	[ju.blé.è.ti]	[ju.bi.lé.èt]
/onibàtâ/ s1-4	[o.ní.bà.tà]	[o.ní.bà.tà]	[o.ní.bà.tà]	[o.ní.bà.tà]			
/gbẹ̀giri/ s1-6	[gbẹ̀.gi.ri]	[gbẹ̀.gi.ri]	[gbẹ̀.gi.ri]	[gbẹ̀.gi.ri]	[gbẹ̀.gri]	[gbẹ̀.gri]	
/flẹ̀ńjò/ s1-6	[fú.lẹ̀.ń.jò]	[flẹ̀.ń.jò]	[fú.lẹ̀.ń.jò]	[flẹ̀.ń.jò]	[flẹ̀.ń.jò]	[flẹ̀.ń.jò]	
/bẹ̀ẹ̀ ní/ s1-4, 7	[bẹ̀.ẹ̀.ni]	[bẹ̀.ẹ̀.ni]	[bẹ̀.ẹ̀.ni]	[bẹ̀.ẹ̀.ni]			[bẹ̀.ẹ̀n]
/osòdi/ s1-4, 7	[osòdi]	[osòdi]	[osòdi]	[osòdi]			[osòòd]
/àwòdi/ s1-4	[àwòdi]	[àwòdi]	[àwòdi]	[àwòdi]			
/sùgbón/ s1-6	[sùgbón]	[sùgbón]	[sùgbón]	[sùgbón]	[sgbón]	[sgbón]	
/àsìgbó/ s1-4	[àsìgbó]	[àsìgbó]	[àsìgbó]	[àsìgbó]			

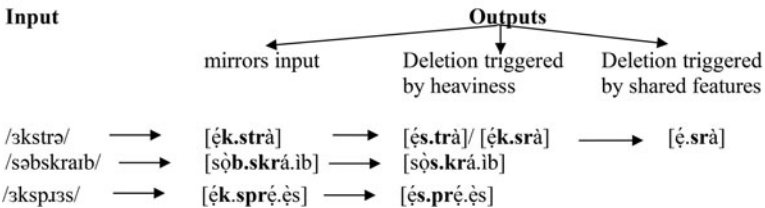
Table 1: Yoruba lexical stratification

(20) Gradient cluster variation

a. Three/two optimal variants for /sprait/, /strai/, and /splæf/ (Stratum 4)



b. Three/two optimal variants for /3kstrə/, /sɔbskraib/, and /3kspɹɪs/



The variation exemplified in (20a,b) relates to how many of the consonants in a cluster in an input is present in the output. The variation exemplified in (20a) is not the full range of possible variants for each of the examples. For instance, the

full range for /straik/ is: /straik/ → [strá.ìk] ~ [srá.ìk] ~ [strá.ìkì] ~ [srá.ìkì]. I have included pairs that differ only with respect to the number of consonants in a cluster, to highlight the point of discussion here. We do not have a way to capture this dimension in the stratification presented so far.

To account for this gradience, I argue that we need a model of variation that can handle all the data presented so far with minimal ad hoc stipulations. But before we consider other models of variation, let us consider some alternatives. We could say that perhaps we have four additional strata in our partial-order model: one which allows CCC clusters and resolves codas, one which allows CCC clusters but preserves codas, one which allows CCCC clusters and resolves codas, and one which allows CCCC clusters but preserves codas. The problem is that it not only unnecessarily and uneconomically duplicates Grammars 2 and 4, but also misses the question of why we have the grammars in the first place. The main distinction between the grammars is whether a cluster and/or a coda is resolved or hypercorrected, and not to what extent a cluster and/or a coda is resolved or hypercorrected, which is what our hypothetical grammars suggest should be the main distinction. This hypothetical path also misses the generalization that, for instance, the difference between examples like *sípráût* and *spráût* or *sípráûtì* and *spráûtì* (Sprite), is strictly a matter of harmonic gradience rather than a categorical one. Instead, we could incorporate Coetzee's (2004) ranked-winners approach so that, for example, it could be suggested that cluster-preserving grammars (mainly Grammars 2 and 4) give room for two candidates to become winners, but only along a single dimension (i.e., the dimension of consonant sequence). This line of investigation is however found to be too ad hoc. In the next section, I consider learning-theoretic models of variation and argue that despite the fact that these models alone cannot account for all the patterns of variation described above for Yoruba, they do not lead to as many ad hoc stipulations. Therefore, of all the models assessed, learning-theoretic OT models prove to be superior to non-probabilistic categorical grammars such as those discussed in sections 4.2, 4.3, and 4.4.

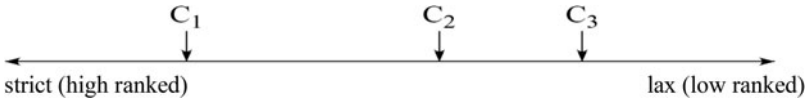
#### 4.5 The Yoruba data and learning-theoretic models (GLA and Maximum Entropy)

Optimality Theory (Prince and Smolensky 2004 [1993]) started out as a non-probabilistic model of language producing categorical grammars. Most of the OT models proposed for handling free variation such as the indexation, co-phonology, and ranked-winners approaches have equally been mainly non-probabilistic in nature. However, a line of research that seeks to address the question of learnability within OT has given rise to learning algorithms that aim to model how grammar is acquired, given a set of well-defined constraints and input data. The first such algorithmic model (Constraint Demotion) was proposed in Tesar and Smolensky (2000), but other ranking algorithms have been proposed in Broihier (1995), Pulleyblank and Turkel (1998), among others. However, stochastic OT models such as the Gradual Learning Algorithm (GLA, Boersma (1997)) and Maximum Entropy (MaxEnt, in Goldwater and Johnson 2003) are a departure from other models in that they are able to handle data with free variation and noisy inputs (Goldwater and Johnson

2003). In this section, I show how a stochastic OT model fits the Yoruba data with fewer ad hoc stipulations, compared to the (non-probabilistic) categorical models discussed in the previous subsections.

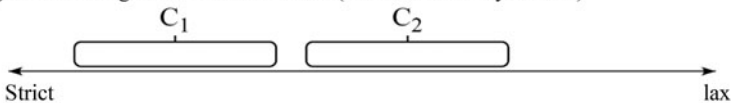
The GLA, first of all, assumes that constraints are arranged on a linear scale (Boersma and Hayes 2001), as graphically represented in (21). A higher value indicates the higher ranking of a constraint, while a lower value shows that a constraint is ranked low.

(21) Constraint ranking on a continuous scale (Boersma and Hayes 2001)

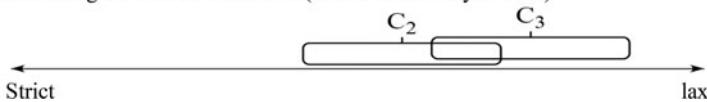


In (21), Constraint  $C_1$  crucially dominates  $C_2$ . This is so because the ranking values for the Constraints  $C_1$ ,  $C_2$ , and  $C_3$  are assumed to be single points on the scale. The GLA proposes instead that the ranking values should be understood as ranges, so that the selection point (within the range) for a given constraint is fixed at evaluation time. When the ranges of two given constraints do not overlap, as in (22a), categorical grammars are produced without any variation. The pattern in (22a), then, is that assumed in Standard OT. On the other hand, free variation results when the ranges overlap, as in (22b). If, for example, the selection value for  $C_2$  is at the leftmost edge of its range and that of  $C_3$  is at the rightmost edge of its range, this will result in the ranking  $C_2 \gg C_3$ , and a candidate that does best on this ranking will become the winner. A converse of this, where the selection value for  $C_3$  is at the leftmost edge of its range and that of  $C_2$  is at the rightmost edge of its range, will result in the ranking  $C_3 \gg C_2$ , and produce a candidate that performs best (given this ranking) as a winner.

(22) a. Categorical ranking on a continuous scale (Boersma and Hayes 2001)



b. Free ranking on a continuous scale (Boersma and Hayes 2001)



The MaxEnt model introduced in Goldwater and Johnson (2003) operates on these same fundamental assumptions. In other words, the algorithms of the GLA and the MaxEnt model are based on these assumptions (see Boersma and Hayes 2001 for more details on the GLA, and Goldwater and Johnson (2003) for the MaxEnt model).<sup>8</sup>

<sup>8</sup>Both of these algorithms have now been programmed into user-friendly applications. The GLA is available in OT Soft (Hayes 2021) while the MaxEnt model is available in the MaxEnt

While both models operate on the fundamental assumptions outlined above, they differ in three respects. First, the MaxEnt model is able to account for cumulative constraint violation while the GLA cannot (Goldwater and Johnson 2003). Second, the MaxEnt model, whose application goes well beyond phonology, appears to be more mathematically motivated than the GLA, which is a ‘somewhat ad hoc model’ developed specifically for learning linguistic constraints (Goldwater and Johnson 2003). These two differences seem to suggest that the MaxEnt model is superior to the GLA. However, the third difference is an area where the GLA performs better: the GLA learns on-line while the MaxEnt model learns off-line (Jager 2007). Because human language works on-line, the GLA provides a model of language acquisition that is closer to how language is actually acquired than the MaxEnt model does (Jager 2007). In the following analysis, I adopt the MaxEnt model available in the MaxEnt Grammar Tool, noting however that this could equally be done with the GLA model in OT Soft, since ‘there is no empirical evidence to favour the one model above the other’ (Jager 2007).

Before beginning the MaxEnt analysis proper, we must update the set of constraints we used in order to accommodate the new pattern of variation illustrated in (20), generalizable in the following way: CCCC clusters are considered to be too heavy in Contemporary Yoruba. For this reason, they are reduced to CCC by deleting the first C in the sequence. This deletion seems to be constrained by the fact that the first C in CCCC sequences is less sonorous than the following C. This idea that the less-sonorous C deletes in a given CC sequence is further reinforced by the reduction of [strá.ìk] to [srá.ìk], where the deleted /t/ is the less sonorous in the sequences /st/ and /tr/. Deletion triggered by shared features is well-documented in the literature. Guy and Boberg (1997: 154), for instance, have observed that deletion is more likely to apply to sequences of segments that have some distinctive features in common. The generalization that in a sequence of two consonants it is the less sonorous consonant that deletes contrasts with the findings in Zec (2007: 194) and Gnanadesikan (2004) where it is shown that it is the more sonorous segment in a CC sequence that deletes. Deletion of a less sonorous segment has also been found, however, as in Alber and Plag (2001: 828) in Sranan. The fact that both of these patterns are found in Daasanach (Nishiguchi 2004) suggests that both of these patterns may be attested even in a single language.

Reduction of CCCC to CCC takes place only through deletion. However, reduction of CCC to CC takes place through both deletion and epenthesis. The question then is: What determines if a CCC cluster will be reduced to CC by epenthesis or by deletion? What can be drawn from the data in (20) is that deletion takes place in CCC clusters only if two adjacent consonants share the same place of articulation. Otherwise, epenthesis takes place. Now consider the following constraints; I assume the sonority hierarchy in (23e).

---

Grammar Tool (Hayes 2009). Both applications learn corpus data with actual frequency. However, since there are no corpora of the Yoruba data that can be used as learning data at this time, I do not consider corpus learning in the analysis presented in this section.

- (23) a. MAX-C/CONTRAST=PLACE (MAX-C/C=P): Deletion of a consonant that does not agree in place of articulation with an adjacent segment is forbidden (See Côté 2000: 170).
- b. MAX-C<sub>>SON</sub>[CC]: for a sequence of two consonants CC present in the input, the most sonorous C is also present in the output.
- c. DEP<sub>[CCCC]</sub>D: Phonological epenthesis is forbidden in the domain of a CCCC cluster.
- d. DEP<sub>[CC/SHARE=PLACE]D</sub> (DEP<sub>[CC/SHARE=P]</sub>): Phonological epenthesis is forbidden in the domain of CC sequences where the two consonants share the same place of articulation.
- e. Sonority Hierarchy: Obstruents (Stops < Fricatives) < Nasals < Liquids (l < r) < Glides < Vowels (highV < lowV) (Cho and King 2003).

All of these constraints are motivated by the descriptions above. (23a) prevents deletion of a segment that does not share any feature with an adjacent segment. (23b) ensures that if deletion must occur, then it must be the less sonorous C that deletes. (23c) is motivated by the generalization discussed above that CCCC clusters are reduced to CCC only through deletion while epenthesis is forbidden. This means that epenthesis is forbidden in the super-heavy (CCCC) cluster domain, even though it is allowed in CCC cluster domain. (23d) also is motivated by the generalization that CCC clusters, where two of the consonants share the same place of articulation, reduce to CC via deletion, while epenthesis is forbidden. Let us first consider the question of why some CCCC sequences reduce to CCC and then to CC, but some others reduce only to CCC as seen in (20).

(24) CCCC cluster reduction

		MAX-C <sub>&gt;SON</sub> [CC]	MAX-C/C=P
I. /səbskraib/	(a) [sò <b>b</b> .skrá.ì.b]		
	(b) [sò <b>.</b> skrá.ìb]		*
	(c) [sò <b>b</b> kráìb]	*!*	**
	(d) [sò <b>bs</b> rá.ìb]		**!
	(e) [sò <b>bs</b> ká.ìb]	*!	*
	(f) [sò <b>sr</b> áìb]		**!*
	(g) [sò <b>sk</b> áìb]	*!	**
	(h) [sò <b>bs</b> áìb]	*!	***
	(i) [sò <b>br</b> áìb]	*!*	*****
	(j) [sò <b>b</b> káìb]	*!***	***
	(k) [sò <b>kr</b> áìb]	*!*	***

II. /ɛkstrim/	☞ (a) [ɛkstriim]		
	☞ (b) [ɛstriim]		*
	(c) [ɛktriim]	*!*	*
	☞ (d) [ɛksriim]		
	(e) [ɛkstiim]	*!	
	(f) [ɛksiim]	*!	**
	☞ (g) [ɛsriim]		*
	(h) [ɛktiim]	*!***	*
	(i) [ɛkriim]	*!*	*
	(j) [ɛstiim]	*!	*
	(k) [ɛtriim]	*!*	**

The tableau in (24) shows that  $MAX-C_{>SON}[CC]$  is highly ranked, meaning a single violation is fatal. All the winning candidates in (24.I) and (24.II) obey it.  $MAX-C/C=P$ , on the other hand, is more flexible in that winners can violate it once. If we were to incorporate Coetzee’s (2004) ranked-winners approach in a partial-order grammar, as mentioned in section 4.4, we would suggest that the critical cut-off point for  $MAX-C/C=P$  is specified to be below the best two evaluations, where no-violation is the best evaluation, and a single violation is second best. If the grammar were so designed, then only candidates that do not violate  $MAX-C_{>SON}[CC]$  and earn at most a single violation on  $MAX-C/C=P$  could be optimal. But the problem lies in defining and constraining the placement of critical cut-off points below the standard ‘best’ evaluation. Besides, incorporating this ranked-winners mechanism in a partial-order grammar already enhanced with the addition of lexical indexation leads to too many ad hoc stipulations, unnecessary in the face of more economical models. Consider in the following tableau why epenthesis does not apply in CCCC sequences and how the repair of CCC clusters is effected by either epenthesis or by deletion.

(25) CCCC and CCC reduction: deletion, and epenthesis

		MAX-C/SON[CC]	MAX-C/C=P	DEP <sub>[CCCC]<sub>D</sub></sub>	DEP <sub>[CC/SHARE=P]</sub>
I. express	☞ (a) [éksprees]				
	☞ (b) [ésprees]		*		
	(c) [éspirees]		*	*!	
	(d) [éksrees]		**!		
	(e) [ékspees]	*!	*		
	(f) [ékssees]	*!	***		
	(g) [éksiprees]			*!	
II. Sprite	☞ (a) [sprait]				
	☞ (b) [siprait]				
	(c) [spirait]				
	(d) [srait]		**!		
	(e) [prait]	*!	*		
III. splash	☞ (a) [splash]				
	☞ (b) [spulaash]				
	☞ (c) [siplaash]				
	(d) [slaash]		**!		
	(e) [plaash]	*!	*		
IV. strike	☞ (a) [straik]				
	☞ (b) [sraik]				
	(c) [sitraik]				*!
	(d) [stiraik]				*!
	(e) [traik]	*!			

The fact that candidates such as (25.I.g) in this grammar are not optimal, due to violating DEP<sub>[CCCC]<sub>D</sub></sub>, captures the generalization that epenthesis is disfavored in CCCC cluster domains. In CCC cluster reduction, two conditions need to be met for deletion to take place: the deleting consonant must share the same place of articulation with adjacent segments (captured by MAX-C/C=P), and the C to be deleted must be the one that is the least sonorous in a CC sequence. Candidates like (25.IV.b), which obey these two constraints, successfully reduce a CCC cluster to a CC cluster via deletion. Candidates like (25.II.d and 25.III.d) are not grammatical due to fatal violations on MAX-C/C=P. Since deletion fails in these two competitions, epenthesis takes place, making Candidates (25.II.b, 25.II.c, and 25.III.b and 25.III.c) optimal. The



epenthetic Candidates (25.IV.c and 25.IV.d) violate  $DEP_{[CC/SHARE=P]}$ . I have gone into these details to illustrate how the constraints used in the MaxEnt analysis that follows work independently of a learning algorithm. With this updated set of constraints, let us now proceed to a MaxEnt analysis. For this analysis, I prepared an input file with 28 input forms. Table 2 shows a MaxEnt grammar learned by the algorithm at  $\mu=0.0$ ,  $\sigma^2=100000.0$ ). All MaxEnt grammars reported here are, by default, learned at these values.

Constraints	Weights
$DEP_{[CCCC]}$	11.79511156
$MAX-C_{>SON}$	11.59946591
$DEP_{[CC/SHARE=P]}$	10.04511483
$MAX-C/C=P$	1.143007328
FAITH	0.722117439
*CCC	0.699652195
NoCODA	0.652604509
*COMPLEX	0.311387169
*[.V#]	0
CC-SEQ	0

**Table 2:** All lexical items in a single lexicon with a single grammar

Since there is no data available to train the MaxEnt algorithm, all winning candidates are specified for a frequency of 1, while all losing candidates are specified for a frequency of 0.<sup>9</sup> The grammar reported in Table 2 is constructed in such a way that for any given input, all the optimal candidates are specified as winners, or in other words, it is an attempt to model all the variation reported above within a single grammar using a single unstratified lexicon. The grammar in Table 2 correctly portrays the generalization that  $MAX-C_{>SON}$ ,  $DEP_{[CCCC]}$ ,  $MAX-C[CC]$  strictly dominate other constraints, even when these given constraints have weights closer to one another (a pattern suggesting non-strict domination and the possibility of re-ranking, especially in the case of  $MAX-C_{>SON}$  and  $DEP_{[CCCC]D}$ ). This grammar also correctly captures the ranking among \*CCC, FAITH, NoCODA, and \*COMPLEX. The fact that \*CCC and FAITH have similar weights suggests that they could be re-ranked at evaluation time to generate optimal forms that satisfy FAITH at the expense of \*CCC, and optimal forms that satisfy \*CCC at the expense of FAITH. Also, even though the ranking among FAITH, NoCODA, and \*COMPLEX holds, the fact that their weights are not far apart suggests that they could be re-ranked at evaluation time in a way that would lead to the generation of the variation related to cluster and coda retention and resolution. However, this grammar fails in two respects: (1) it incorrectly suggests that the hypercorrection

<sup>9</sup>It is hoped that when the corpus of contact-induced changes being compiled by the author is ready, more could be done in terms of matching the frequency generated by the MaxEnt model with real-life data.

constraints, \*[V#] and CC-SEQ, are inactive given that they are judged to have a weight of 0, contrary to what we have empirically established; (2) this grammar over-generates. Consider Table 3 for concreteness:

Input	Candidate	Winner	p-value
Sugbon	sùgbón	1	0.588633
	sgbón	1	0.210116
	*sùgbóns	0	0.14831
	*sgbóns	0	0.05294
Asigbo	àsìgbó	1	0.588633
	*àsgbó	0	0.210116
	*àsìgbóós	0	0.14831
	*àsgbóós	0	0.05294
iya-koko	ìyá-kókó	1	0.588633
	ìyá-kókós	1	0.14831
	*ìykókó	0	0.210116
	*ìykókós	0	0.05294

**Table 3:** Candidates’ probability of occurrence

Table 3 shows the probabilities generated by the MaxEnt grammar in Table 2 for some of the outputs. The grammar suggests that faithful candidates like *sugbon*, devoid of clusters and codas, will occur at chance 59% of the time, cluster hypercorrection forms 21% of the time, and coda hypercorrection 14% of the time. While this pattern might be accurate for inputs that simultaneously have faithful, cluster hypercorrection, and coda hypercorrection optimal forms, no input as yet has all of these optimal forms. An input either has faithful and cluster hypercorrection optimal forms, or faithful and coda hypercorrection optimal forms. This means that we have no way yet to explain the following: a) why faithful and cluster hypercorrection candidates are optimal for *sùgbón* but not for *àsìgbó*; b) why coda hypercorrection forms are not optimal for all of these inputs; c) why only faithful candidates are optimal for *àsìgbó*—as it stands, this grammar incorrectly rates *sgbón*, and *àsgbó* as grammatical; and d) why only faithful and coda hypercorrection forms are optimal for an input like *ìyá-kókó*. Without further mechanism, this grammar inappropriately over-generates unattested forms like *àsgbó*.

This problem of over-generation was noted by Moore-Cantwell and Pater (2016) who observe that the MaxEnt model in its unelaborated form is appropriate for (free) variation but not lexically-conditioned variation (or exceptionality). The main obstacle encountered by the grammar reported in Table 2 lies in the fact that it cannot handle lexically-specific constraints which do not hold of all the lexical items in the Contemporary Yoruba lexicon. See Anttila and Magri (2018) for further discussion on how the MaxEnt model overgenerates. To avoid the problem of over-generation, I incorporate lexical indexation into the MaxEnt model reported in Table 2, as in Moore-Cantwell and Pater (2016). This way, the Contemporary

Yoruba lexicon is characterized by stratification, where there is the Core Stratum (henceforth CS) comprising all lexical items (LIs) in Yoruba, whether native or loan; the Cluster Hypercorrection Periphery Stratum (henceforth CIS) comprising only a set of those LIs that permit cluster hypercorrection; and Coda Hypercorrection Periphery Stratum (henceforth CoS), comprising only those LIs that are subject to  $*[.V\#]$ . It can be proposed that at evaluation time, speakers have access to the information that CC-SEQ and  $*[.V\#]$  are indexed respectively for the CIS and CoS so that they are more appropriately represented as  $CC-SEQ_{CIS}$  and  $*[.V\#]_{CoS}$ . During evaluation, the grammar does not assess any of the candidates for non-hypercorrection inputs like *àsípa* and *àsìgbó* against the lexically-indexed hypercorrection constraints,  $CC-SEQ_{CIS}$  and  $*[.V\#]_{CoS}$ ; this is because these inputs are not indexed for CIS and CoS, but belong only to CS. Candidates generated for cluster hypercorrection inputs like *sùgbón* and *popular* are assessed against  $CC-SEQ_{CIS}$  in addition to the rest of the constraints because they are indexed for it. They are not assessed by  $*[.V\#]_{CoS}$  because they are not indexed for it. Likewise, candidates generated for coda hypercorrection inputs like *şàgámù* are assessed by  $*[.V\#]_{CoS}$  for the same reason that they are indexed for it and not by  $CC-SEQ_{CIS}$  because they are not indexed for it. This kind of MaxEnt evaluation that is constrained by lexical indexation can be represented as in (26):

(26) A lexically-indexed MaxEnt grammar

		DEP <sub>[CCCCP]</sub>	MAX-C-SON[CC]	DEP <sub>[CCSHARE-P]</sub>	$*[.V\#]_{CoS}$	CC-SEQ <sub>CIS</sub>	FAITH	NOCODA	*COMPLEX	MAX-C/C=P	*CCC	<i>p</i>
		11.57	11.52	9.40	3.14	2.78	1.60	1.52	1.29	1.19	0.44	
I. <i>àsìgbó</i>	☞ (a) <i>àsìgbó</i>											.91
	(b) <i>àsgbó</i>						1		1			.05
	(c) <i>àsìgbóós</i>						1	1				.04
II. <i>sùgbón<sub>CIS</sub></i>	☞ (a) <i>pópùlà</i>					1						.51
	(b) <i>póplà</i>						1		1			.46
	(c) <i>pópùlààs</i>					1	1	1				.02
IV. <i>şàgámù<sub>CoS</sub></i>	☞ (a) <i>şàgámù</i>				1							.48
	(b) <i>şgámù</i>				1		1		1			.03
	(c) <i>şàgáám</i>						1	1				.49

In (26), the row immediately below the constraints indicates the weights learned by the MaxEnt model for each of the constraints, which determine the hierarchy of the constraints. As mentioned previously, the fact that the weights for some of the constraints are close to one another means that they can be re-ranked at evaluation

time to produce the two kinds of regular variation manifested in Contemporary Yoruba Phonology: (a) categorical variation: BÚRÉDÌ, BRÉDÌ, BÚRÉÈD and BRÉÈD; and (b) gradient variation: ÈKSTRÀ, ÈSTRÀ, ÈKSRÀ, and ÈSRÀ. The last column, labeled *p*, indicates the frequency probability generated by the model for each of the relevant candidates. For example, the model predicts that Candidate (26.I.a) will be optimal 91% of the time, as opposed to the other candidates whose probabilities of occurrence are very low.

This MaxEnt analysis is in line with the claim in section 4.4 that the ongoing changes in Yoruba actively take place at the periphery, where it is possible that more and more LIs may be indexed to any of the periphery strata, until the two periphery constraints become part of the core grammar of the language. In the various analyses discussed in this article, we have seen that none of the approaches examined is able to independently account for all the categorical, gradient, and lexically-conditioned variation that characterizes Contemporary Yoruba. However, of all the variation approaches assessed, the MaxEnt model appears to be the most economical and the most successful in that we only need to augment it with lexical indexation for it to be able to capture the full extent of the variation in the Yoruba data. The analysis is also more elegant in another respect: rather than postulating a proliferation of seven grammars, the lexically-indexed MaxEnt model generates only three sub-grammars: the Core grammar which all lexical items are subject to, and two peripheral grammars which only a select number of lexical items are subject to. This model also helps us to situate the core–periphery structure of the Yoruba lexicon within the three patterns of lexical stratification documented in Hsu and Jesney (2018). The lexically indexed MaxEnt model in (26) reveals that there are two independent supersets at the periphery of the Yoruba phonological lexicon: a) one which contains all LIs in the Core (subject to the Core constraints) and those that are subject to the Periphery constraint CC-SEQ<sub>CIS</sub> and b) another one that equally contains all LIs in the Core (subject to the Core constraints) and all those lexical items subject to the Periphery constraint \*[V#]<sub>CoS</sub>. It could be concluded that Yoruba provides evidence for a core–periphery pattern that might be termed ‘double supersets at periphery’, where two peripheral strata exhibiting the ‘superset at periphery’ pattern are characterized by phonological generalizations that hold independently for each stratum.

## 5. PREDICTIONS OF THE ANALYSES

One important prediction that I have made in my analysis is that if the variation seen is indeed a case of change in progress, then it follows that the final stage of the change (Period 3b) will not be characterized by constraint re-ranking or further stratification, but rather by changes in lexical indexation, so that more and more lexical items are indexed for CC-SEQ<sub>CIS</sub> and \*[V#]<sub>CoS</sub> until these two constraints become part of the core grammar, and many new lexical items emerge with invented clusters and codas. To be sure, there is only one more grammar that can be admitted into the Yoruba grammar that we have established, typologically speaking, involving coda hypercorrection and cluster hypercorrection simultaneously and thus only surfacing at the Periphery. It follows then that the eventual

change will be in terms of more words being indexed to the existing grammars at the Periphery. The culmination of the change will be a point in time where hypercorrection holds of all lexical items in the lexicon with the right phonological context. At such a point, forms like *\*awood* ‘hawk’ and *\*àsgbò* ‘mishearing’ with the right condition for hypercorrection will be grammatical. As a result of this, our analysis also makes a prediction that change in progress must be understood in terms of a core–periphery structure, where ongoing change actively takes place at the periphery spreading through the entire lexicon one at a time until the periphery becomes part of the core so that grammatical innovations that hold of some words at the periphery gradually spread until they become a regular part of the core of a language’s grammar. This way, the general assumption in the sociolinguistic literature (e.g., Aitchison 2004: 4) that change takes place gradually is put in theoretical perspective.

## 6. CONCLUSION

The following generalizations have emerged from the foregoing discussion on Contemporary Yoruba phonology: a) Yoruba is in the second stage of a diachronic change, where variation is the major hallmark of its GRAMMAR; b) this resulting variation is either a stable variation which will continue for a long time to come, or a change in progress that will eventually culminate in the lexically unconstrained permission of clusters and codas; c) if this variation is an instance of change in progress, then the final stage is going to be characterized mainly by changes in lexical indexation, rather than further constraint re-ranking; d) the different patterns of cluster and coda realization described in section 3 are not random, but are a well-organized stratification in a phonological lexicon; e) this stratification is organized in a core–periphery fashion where the Core is characterized by free and gradient variation and the Periphery is characterized by lexically-conditioned variation; f) ranked-winners, indexed constraint, partial-order co-phonology, and MaxEnt approaches to variation and lexical stratification are inadequate to independently account for the variation now observable in Yoruba; g) however, a MaxEnt model augmented with lexical indexation provides the most successful and parsimonious account of the Yoruba data. Overall, more research is needed to understand fully the contact-induced phenomena described in this article. Future work would do well to shed light on how the seven grammars identified for Yoruba are acquired by children. It is furthermore the hope that this article will inspire similar investigations in other African languages with similar prolonged contact with a former colonial language.

## REFERENCES

- Adebayo, Taofeeq. 2022. Contact-induced variation and change in Yoruba: Optimality in grammar and its social dimensions. Doctoral dissertation, Tulane University.

- Adeniyi, Kōlawōle. 2015. Hybrid sounds in Yoruba. *Ihafa: A Journal of African Studies* 7 (1): 1–20.
- Aitchison, Jean. 2004. *Language change: Progress or decay?* Cambridge: Cambridge University Press.
- Akinlabi, Akinbiyi. 2004. The sound system of Yoruba. In *Understanding Yoruba life and culture*, ed. Nike Lawal, Mathew Sadiku, and Ade Dopamu, 453–468. Trenton, NJ: Africa World Press.
- Alaroye TV. 2021. Tinubu n sun lowo, e ma distoqobu re. Retrieved July 29, 2021. <<https://www.facebook.com/alaroyeOnlinetv/posts/4241955449205202>>
- Alber, Birgit, and Ingo Plag. 2001. Epenthesis, deletion and the emergence of the optimal syllable in Creole: The case of Sranan. *Lingua* 111(11): 811–840.
- Anttila, Arto. 1997. Deriving variation from grammar. In *Variation, change, and phonological theory*, ed. Frans Hinskens, Roeland van Hout, and W. Leo Wetzels, 35–68. Amsterdam: John Benjamins.
- Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20(1): 1–42.
- Anttila, Arto. 2007. Variation and optionality. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 519–536. Cambridge: Cambridge University Press.
- Anttila, Arto. 2009. Derived environment effects in colloquial Helsinki Finnish. In *The nature of the word: Essays in honor of Paul Kiparsky*, ed. Kristin Hanson and Sharon Inkelas, 433–60. Cambridge, MA: MIT Press.
- Anttila, Arto, and Young-mee Yu Cho. 1998. Variation and change in Optimality Theory. *Lingua* 104(1–2): 31–36.
- Anttila, Arto, and Giorgio Magri. 2018. Does MaxEnt overgenerate? Implicational universals in maximum entropy grammar. In *Proceedings of the Annual Meetings on Phonology*, vol. 5.
- Badmus, Issa Abdulwaheed. 2020. Emergence of Arabic new loan words among Yoruba speakers in Southwestern Nigeria: A critical analysis. *Journal of Management and Social Sciences* 9(1): 829–42.
- Bamgbose, Ayo. 1982. Yoruba orthography. In *Orthographies of Nigerian languages. Manual I*, ed. Ayo Bamgbose. Lagos: National Language Centre, Ministry of Education.
- Blevins, Juliette. 2006. Syllable typology. *Encyclopedia of Language and Linguistics, Vol. 12 [Spe - Top]*: 333–37. Oxford: Elsevier.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the Univ. of Amsterdam*, volume 21, 43–58.
- Broihier, Kevin. 1995. Optimality theoretic rankings with tied constraints: Slavic relatives, resumptive pronouns and learnability. Ms. MIT. Rutgers Optimality Archive ROA-46. <<http://ruccs.rutgers.edu/roa.html>>
- Chappell, Whitney. 2014. Reanalysis and hypercorrection among extreme /s/ reducers. *University of Pennsylvania Working Papers in Linguistics* 20(2).
- Cho, Young-mee Yu, and Tracy Holloway King. 2003. Semi-syllables and universal syllabification. In *The syllable in Optimality Theory*, ed. Caroline Féry and Ruben van de Vijver. Cambridge: Cambridge University Press.
- Chung, Karen Steffen. 2006. Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication* 16(2): 197–214. doi: 10.1075/japc.16.2.04chu
- CMS. 1913. *A dictionary of Yoruba language*. Lagos: Church Missionary Society.
- Coetzee, Andries. 2004. What it means to be a loser: Non-optimal candidates in Optimality Theory. Doctoral dissertation, University of Massachusetts Amherst.
- Côté, Marie-Hélène. 2000. Consonant cluster phonotactics: A perceptual approach. Doctoral dissertation, MIT. Available as ROA-548 from the Rutgers Optimality Archive.

- Flack, Kathryn. 2009. Constraints on onsets and codas of words and phrases. *Phonology* 26(2): 269–302.
- Fukazawa, Haruka 1998. Multiple input-output faithfulness relations in Japanese. Rutgers Optimality Archive, ROAa260–0598.
- Gess, Randall. 1996. Optimality Theory in the historical phonology of French. Doctoral dissertation, University of Washington.
- Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In *Fixing priorities: Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 73–108. Cambridge: Cambridge University Press.
- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*, 111–120.
- Gut, Ulrike B. 2008. Nigerian English: Phonology. In *Varieties of English 4: Africa, South, and Southeast Asia*, vol. 4, 35–54. Berlin: Mouton de Gruyter.
- Guy, Gregory R., and Charles Boberg. 1997. Inherent variability and the Obligatory Contour Principle. *Language Variation and Change* 9(2): 149–164.
- Hayes, Bruce. 2009. Manual for Maxent Grammar Tool, UCLA. <<https://linguistics.ucla.edu/people/hayes/MaxentGrammarTool/ManualForMaxentGrammarTool>>
- Hayes, Bruce. 2021. OTSoft: Constraint Ranking Software – Version 2.5. Manuscript, UCLA. <<http://www.linguistics.ucla.edu/people/hayes/otsoft/OTSoftManual.doc>>
- Holden, Kyril. 1976. Assimilation rates of borrowings and phonological productivity. *Language* 52(1): 131–147.
- Holt, D. Eric. 1997. The role of the listener in the historical phonology of Spanish and Portuguese: An Optimality-Theoretic account. Doctoral dissertation, Georgetown University.
- Hsu, Brian, and Karen Jesney. 2018. Weighted scalar constraints capture the typology of loanword adaptation. In *Proceedings of the Annual Meetings on Phonology*, vol. 5.
- Hutton, John. 1996. Optimality Theory and historical language change. *The 4th Phonology Workshop*, University of Manchester.
- Inkelas, Sharon. 1998. The theoretical status of morphologically conditioned phonology: A case study of dominance effects. In *Yearbook of morphology*, ed. Gert Booij and Jaap van Marle. Dordrecht: Kluwer Academic Publishers.
- Inkelas, Sharon and Cheryl Zoll. 2007. Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45(1): 133–171.
- Itô, Junko, and Armin Mester. 1995a. The core-periphery structure in the lexicon and constraints on re-ranking. In *Papers in Optimality Theory*, ed. Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, 181–210. Amherst, MA: GLSA, University of Massachusetts.
- Itô, Junko, and Armin Mester. 1995b. Japanese phonology. In *The handbook of phonological theory*, ed. John Goldsmith, 817–38. Oxford: Blackwell.
- Itô, Junko, and Armin Mester. 1999. The phonological lexicon. In *The Handbook of Japanese Linguistics*, ed. N. Tsujimura, 62–100. Oxford: Blackwell.
- Itô, Junko, and Armin, Mester. 2003. Systemic markedness and faithfulness. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 39(1), 665–689.
- Jacobs, Haike. 1995. Optimality Theory and sound change. In *Proceedings of the North-Eastern Linguistic Society* 25 (2), ed. Jill Beckman, 219–232. Amherst, MA: GLSA Publications.
- Jacobs, Haike. 1996. Lenition and Optimality Theory. In *Aspects of Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages*, XXIV, ed.

- Claudia Parodi, Carlos Quicoli, Mario Saltarelli, and Maria Luisa Zubizarreta, 253–265. Washington, DC: Georgetown University Press.
- Jager, Gerhard. 2007. Maximum entropy models and Stochastic Optimality Theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. J. Grimshaw, J. Maling, C. Manning, J. Simpson, and A. Zaenen, 467–479. Stanford, CA: CSLI.
- Janda, Richard D., and Julie Auger. 1992. Quantitative evidence, qualitative hypercorrection, sociolinguistic variables—and French speakers' headaches with English h/Ø. *Language and Communication* 12(3–4): 195–236.
- Kang, Yoonjung, Michael Kenstowicz, and Chiyuki Ito. 2008. Hybrid loans: A study of English loanwords transmitted to Korean via Japanese. *Journal of East Asian Linguistics* 17(4): 299–316.
- Kimpa, Wendell. 2011. Locality and globality in phonological variation. *Natural Language and Linguistic Theory*, vol. 29(2): 423–465.
- Kiparsky, Paul. 1993. Variable rules. Handout, Rutgers Optimality Workshop 1.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3): 199–244.
- Labov, William. 1966. Hypercorrection by the lower middle class as a factor in linguistic change. In *Sociolinguistics*, ed. William Bright, 84–113. The Hague: Mouton.
- McCarthy, John J. 2007. What is Optimality Theory? *Linguistics Department Faculty Publication Series*. Paper 93. <[http://scholarworks.umass.edu/linguist\\_faculty\\_pubs/93/](http://scholarworks.umass.edu/linguist_faculty_pubs/93/)>
- McCarthy, John J. 2008. *Doing Optimality Theory: Applying theory to data*. Malden, MA: Blackwell Publishing.
- McCarthy, John, and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology (YOMO)*, 79–153.
- McCarthy, John J., and Prince, Alan. 1995. Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249–384. Amherst, MA: GLSA Publications.
- McCarthy, John J., and Prince, Alan. 1999. Faithfulness and identity in prosodic morphology. In *The Prosody–Morphology Interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, 218–309. Cambridge: Cambridge University Press.
- Mendoza-Denton, Norma Catalina. 1997. *Chicana/Mexicana identity and linguistic variation: An ethnographic and sociolinguistic study of gang affiliation in an urban high school*. Doctoral dissertation, Stanford University.
- Miglio, Viola, and Bruce Moren. 2003. Merger avoidance and lexical reconstruction: An OT model of the Great Vowel Shift. In *Optimality Theory and Language Change*, ed. D. Eric Holt, 191–228. Dordrecht: Kluwer Academic Publishers.
- Moore-Cantwell, Claire, and Joe Pater. 2016. Gradient exceptionality in Maximum Entropy Grammar with lexically specific constraints. *Catalan Journal of Linguistics* 15: 53–66.
- Nishiguchi, Sumiyo. 2004. Consonant assimilation and sonority: A case study in Daasanach. *Osaka University Papers in English Linguistics*: 39–51.
- Nouveau, Dominique. 1994. Language acquisition, metrical theory, and optimality: A study of Dutch word stress. Doctoral Dissertation, Utrecht University.
- Ola, Olanike. 1995. Optimality in Benue-Congo prosodic phonology and morphology. Doctoral dissertation, University of British Columbia.
- Orgun, Cemil Orhan. 1996. Sign-based morphology and phonology with special attention to Optimality Theory. Ph.D. dissertation, University of California, Berkeley.
- Pater, Joe. 1994. Against the Underlying Specification of an 'Exceptional' English Stress Pattern. In *Toronto Working Papers in Linguistics* 13(1): 95–121.



- Pater, Joe. 2000. Nonuniformity in English stress: the role of ranked and lexically specific constraints. *Phonology* 17(2): 237–274.
- Pater, Joe. 2004. Exceptions in Optimality Theory: Typology and learnability. A talk given at Conference on Redefining Elicitation: Novel Data in Phonological Theory. New York University.
- Pater, Joe. 2007. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*, ed. Leah Bateman, Michael O’Keefe, Ehren Reilly, and Adam Werle 259–296.
- Prince, Alan, and Paul Smolensky. 2004 [1993]. *Optimality Theory: Constraint interaction in Generative Grammar*. Oxford: Blackwell Publishing.
- Pulleyblank, Douglas. 2008. Yoruba vowel patterns: Symmetries through phonological competition. In *Explorations into Language Use in Africa*, ed. Simo Bobda, 125–157. Berlin: Peter Lang.
- Pulleyblank, Douglas, and William J. Turkel. 1998. The logical problem of language acquisition in Optimality Theory. In *Is the best good enough? Optimality and competition in syntax*, ed. Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky, 399–420. Cambridge, MA: MIT Press.
- Steriade, Donca. 2009. The phonology of perceptibility effects: The p-map and its consequences for constraint organization. In *The nature of the word: Essays in honor of Paul Kiparsky*, ed. Kristin Hanson and Sharon Inkelas, 151–180. Cambridge, MA: MIT Press.
- Taiwo, Rotimi. 2009. The functions of English in Nigeria from the earliest times to the present day. *English Today* 25(2): 3–10.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tesar, Bruce. 2004. Computing optimal forms in Optimality Theory: Basic syllabification. In *Optimality Theory in Phonology: A Reader*, ed. John J. McCarthy, 99–117. Malden, MA: Blackwell.
- Ufomata, Titilayo. 1991. ‘Englishization of Yoruba phonology’. *World Englishes* 10(1): 33–51.
- Winford, Donald. 1978. Phonological hypercorrection in the process of decreolization – the case of Trinidadian English. *Journal of Linguistics* 14(2): 277–91. doi: 10.1017/S0022226700005909
- Zamma, Hideki. 2005. Predicting varieties: Partial orderings in English stress assignment. Ms., Kobe City University of Foreign Studies/University College London.
- Zec, Draga. 2007. The syllable. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 161–194. Cambridge: Cambridge University Press.

## APPENDICES

### APPENDIX A: DATA COLLECTION

Three methods were employed in the collection of the data reported in this paper: sociolinguistic interviews, participant observation, and grammaticality judgment. In the following paragraphs, I briefly describe each of these methods.

I conducted 120 sociolinguistic interviews at five different locations in the Yoruba-speaking Southwestern part of Nigeria. I covered two states, Oyo and Lagos. In Oyo State, I interviewed 36 participants in the capital city (Ibadan), 24 in a town (Iseyin), and 12 in a village (Gbonkan village). In Lagos State, I interviewed 36 participants in what I define as the Lagos Megacity (which is a conglomerate of erstwhile towns that have now merged into

Words	Male					Female				
	Coda hypercorrection									
	W	B	Wa	B	B	<i>Iya</i>	<i>Ant</i>	<i>Iya</i>	<i>Iya</i>	<i>Su</i>
				Eja	Ibeji	<i>S</i>	<i>R</i>	<i>Ish</i>	<i>Rod</i>	
om (omú)	5	2	3	5	5	1	4	1	1	3
àwòòd (àwòòdì)	1	2	2	1	3	1	1	1	2	3
Oshòòd (Oshòòdì)	5	4	5	4	5	1	5	1	5	5
ishágààm (ishágààmù)	5	5	5	5	5	4	5	4	4	5
àkààm (àkààmù)	3	3	2	5	5	2	5	1	1	2
àm (àmù)	2	3	5	5	5	4	5	1	1	5
ìwààd (ìwàdíf)	1	2	1	3	2	1	2	1	3	2
ìlààd (ìlààdí)	3	1	1	2	3	1	2	1	1	2
ìbààd (ìbàdí)	1	1	3	1	1	1	2	1	2	2
àán (àánú)	4	1	3	4	5	4	3	1	2	5
aláán (aláánú)	1	4	2	4	3	1	5	1	3	4
ìsaas (ìsaasùn)	1	1	5	3	1	1	5	1	2	5
ìgbaàd (ìgbádùn)	2	3	4	3	3	1	5	1	5	5
agfíít (agfíítí)	1	4	5	1	5	1	5	1	4	4
ìgbàán (ìgbànú)	1	3	1	4	3	1	4	1	5	2
agfíd (agfídí)	3	3	1	3	3	1	2	1	3	2
méjèèj (méjèèjì)	5	5	5	5	5	4	4	5	5	5
àjèèj (àjèjì)	3	4	5	5	5	1	5	1	5	5
òlàáj (òlàájú)	1	1	5	4	3	1	5	1	5	5
àjànàák (àjànàákú)	1	2	2	3	1	1	2	1	2	2
àkífèsí (àkífèès)	3	1	5	5	1	1	4	1	5	5
asíwááj (asíwáájú)	4	4	5	5	4	1	4	1	5	5
	Cluster hypercorrection									
sgbón (sùgbò)	5	5	4	4	3	1	3	1	1	5
àsgbò (àsìgbò)	1	2	3	2	3	1	3	1	2	3
brúkú (burúkú)	5	5	4	4	4	4	4	5	5	5
àskò (àsìkò)	5	5	5	5	4	4	4	4	5	5
gbràà (gbàrà)	5	5	4	4	3	5	5	5	5	3
bèské (bèsúké)	5	5	5	5	5	4	5	5	5	5
ẹ kròlẹ (ẹ kúròlẹ)	5	5	5	5	5	5	5	4	5	5
jubleet (jubilate)	4	5	5	4	5	4	5	5	5	4
asrí (àsírí)	4	3	5	3	1	1	5	1	5	5
préèd (parade)	1	2	1	2	3	2	1	3	2	1
Orísírí (orísírísí)	4	1	5	5	1	1	4	1	1	4
àkrò (àkùrò)	1	3	4	1	1	1	4	1	5	5
bòskòrò (bòsìkòrò)	5	5	5	5	5	5	5	5	5	5
bòskónà (bòsìkónà)	5	5	5	5	5	5	5	5	5	4
abúlé (abulé)	4	4	5	5	3	1	5	1	5	5
Abékta (Abékuta)	4	5	5	5	5	1	5	1	5	5
traka (tìraka)	4	5	3	3	3	1	5	1	5	4
Flàní (Fúlàní)	4	5	5	5	5	1	5	5	5	5

Table 4: (cont.)

**Table 4:** (Cont.)

Words	Male					Female				
tiketke (tiketike)	4	5	4	5	1	4	5	5	5	4
àblà (àbùlà)	5	5	5	5	5	4	5	5	5	5
ablé (abúlé)	5	5	5	5	5	4	4	1	5	5
àsgbé (àsìgbé)	4	3	3	5	3	1	4	2	4	5
Ablèḗgbá (Abúlèḗgbá)	4	5	5	5	5	5	5	1	5	5
rúsdá (rúsídá)	4	5	2	5	1	1	5	1	5	5
àbrò (àbùrò)	5	5	4	5	4	1	5	5	5	5
ablèsowó (abulèsowó)	5	5	3	5	5	1	4	5	5	5
atro (atiro)	3	1	2	3	1	1	2	1	2	3
atra (atura)	1	1	3	1	1	1	1	1	1	3
<b>Gradient cluster retention</b>										
Sípráit (Sprite)	5	5	5	5	3	4	5	5	5	3
Spiráit (Sprite)	1	2	3	3	3	1	5	1	4	3
èkstrà (extra)	4	5	5	5	4	4	5	3	3	5
èstrà (extra)	4	5	5	5	3	4	5	4	4	5
èsrà (extra)	4	5	5	5	5	4	5	5	5	5
èksrà (extra)	4	5	5	5	4	4	5	5	5	5
sòskràìb (subscribe)	5	5	5	5	5	4	5	5	5	5
sòbùskràìb (subscribe)	2	3	3	2	1	1	2	1	2	3
spúláàsh (splash)	3	1	4	3	2	3	1	5	5	5
sípláàsh (splash)	3	1	4	3	3	3	1	5	4	4

**Table 4:** Grammaticality judgment

a huge megacity) and 12 in the village called Oke Agbo, located around the Gberigbe area of Ikorodu. I also conducted participant observations in these locations as well as in Abeokuta, Ogun State. My participant observation was structured in such a way that I moved around with individuals and groups of individuals engaged in activities like watching soccer, working in a workshop, selling wares in a shop or at a market, etc. While doing this, I documented their speech behaviours that were relevant to my research design. This took place between May 2019 and November 2020. The data reported in the main body of the paper and in Appendix B were acquired through these two methods. However, since sociolinguistic interviews and participant observation can yield only positive evidence, which is not enough to answer the question of which form is acceptable and which is not (a question especially relevant for the analysis of lexically conditioned variation at the periphery of the Yoruba GRAMMAR), I complemented these methods with a grammaticality judgment task, part of which is reported in Table 4.

Table 4 is structured in the following way. The first column contains hypercorrected forms with their respective original words in parenthesis. The third row contains the pseudonyms for

the ten participants that were drawn from Iseyin specifically for this task. The participants include five men and five women all between the ages of 18 and 40. They responded to a Likert scale which ranges from 5 to 1, where 5 is interpreted as ‘this form is very grammatical’, 4 as ‘this form is grammatical’, 3 as ‘I doubt if this form is grammatical’, 2 as ‘this form is not grammatical’, and 1 as ‘this form is not grammatical at all’. To adjudge a form as grammatical in a speaker’s grammar, I take 4 to be the benchmark for grammaticality, meaning that if a form is rated 4 or 5 by a speaker, then that form is grammatical in their grammar. If it is rated 3 or below, then the form is not grammatical in the respondent’s grammar.

Three different generalizations, which provide the basis for the discussion in the main text of the paper, arise from this table: (a) some forms like *bòskòrò*, *běské*, *àskò*, *bòskònà*, *şagaam*, and *mějèj* whose rating ranges only between 4 and 5 are already well established in the Contemporary Yoruba grammar; (b) forms like *iláàd*, *ibáàd*, *agíùd*, *àjànààk*, *àsgbó*, *atro*, *atra*, with the right context for hypercorrection are yet to be established in the grammar given that their ratings do not go beyond 3; and (c) forms like *orisrisi*, *aběkta*, *traka*, *am*, *ajeej*, *akaam*, and *olaaj* are intermediate between being totally grammatical or totally forbidden in everyone’s grammar. If this population is expanded to include people of other age groups and people from other locations, it could well turn out that the intermediate pattern is the main fashion in which this hypercorrection variation is organized: lexical conditioning of hypercorrection is different from one individual to another. This would mean that the lexicon for Contemporary Yoruba speakers is not organized uniformly. The lexically indexed MaxEnt model proposed in the main text accounts for this variation in a straightforward way.

## APPENDIX B: THE SEVEN GRAMMARS OF CONTEMPORARY YORUBA PHONOLOGY<sup>10</sup>

<sup>10</sup>For a guide on how to interpret the following tables, see the definitions in (16). Final nasals and prenasal nasals in output forms like *pònkriyòòn* can indicate that the previous vowel is nasal, in line with Yoruba orthography; so, /n/ and /m/ are not phonemic in these cases. In other cases, like *sobontóòn*, the final /n/ is phonemic, meaning that it is a word-final coda. Prenasal nasals can be moraic and occupy a syllable on their own as in [tjé.ń.prá.ri]. They can also be phonemic as in *bank*, where /nk/ is a cluster.

## (27) Core lexicon (Grammars 1, 2, 3, and 4)

	Input	Output				Gloss
		Grammar 1	Grammar 2	Grammar 3	Grammar 4	
<b>Yoruba forms</b>						
1	brégèèd	búrégèèdi	brégèèdi	búrégèèd	brégèèd	be elaborate
2	gbègìrì	gbègìrì	gbègìrì	gbègìrì	gbègìrì	soup made from beans
3	pakuuru mọ	pakuuru mọ	pakuuru mọ	pakuuru mọ	pakuuru mọ	pounce on him/her
4	òfurufú	òfurufú	òfurufú	òfurufú	òfurufú	empty/sky
5	burúkú	burúkú	burúkú	burúkú	burúkú	bad
6	fírì	fírì	fírì	fírì	fírì	sloppy
7	bọ sí kọrọ	bọ sí kọrọ	bọ sí kọrọ	bọ sí kọrọ	bọ sí kọrọ	second-hand cloth
8	bọ sí kọnà	bọ sí kọnà	bọ sí kọnà	bọ sí kọnà	bọ sí kọnà	second-hand cloth
9	alàgbà	alàgbà	alàgbà	alàgbà	alàgbà	old fellow
10	olówó	olówó	olówó	olówó	olówó	rich person
11	alákọwé	alákọwé	alákọwé	alákọwé	alákọwé	educated fellow
12	tálàkà	tálàkà	tálàkà	tálàkà	tálàkà	poor person
13	pariwo	pariwo	pariwo	pariwo	pariwo	shout
14	flẹ́njọ	fúlẹ́njọ	flẹ́njọ	fúlẹ́njọ	flẹ́njọ	enjoy
15	kosté	kosité	kosté	kosité	kosté	a kind of bra
16	skelewú	sikelewú	skelewú	sikelewú	skelewú	a kind of dance
17	yapayáskì	yapayásikì	yapayáskì	yapayásikì	yapayáskì	be.plenty
18	bẹ̀skẹ	bẹ̀sikẹ	bẹ̀skẹ	bẹ̀sikẹ	bẹ̀skẹ	soybean cake
19	skin	sikin	skin	sikin	skin	become as expected
20	roboskẹ	robosikẹ	roboskẹ	robosikẹ	roboskẹ	slang
21	yẹ̀eske	yẹ̀esike	yẹ̀eske	yẹ̀esike	yẹ̀eske	of course
22	yaboskáín	yabosikáín	yaboskáín	yabosikáín	yaboskáín	trendy
23	skẹ̀ntẹ̀lẹ̀	sikẹ̀ntẹ̀lẹ̀	skẹ̀ntẹ̀lẹ̀	sikẹ̀ntẹ̀lẹ̀	skẹ̀ntẹ̀lẹ̀	spread (of headscarf)
24	skontolo	sikontolo	skontolo	sikontolo	skontolo	elegant (of headscarf)
25	skélébé	sikélébé	skélébé	sikélébé	skélébé	small
26	skólóbó	sikólóbó	skólóbó	sikólóbó	skólóbó	small

27	askarí	asikarí	askarí	asikarí	askarí	police officer
28	wòós	wòósí	wòósí	wòós	wòós	hey
29	gbaas-gbòós	gbaasi-gbòósí	gbaasi-gbòósí	gbaas-gbòós	gbaas-gbòós	fracas
30	twálè	túálè	twálè	túálè	twálè	a salutation
31	skònskòṅ	sikònskòṅ	skònskòṅ	síkònskòṅ	skònskòṅ	madness
32	ponkriyọṅ	ponkiriyoṅ	ponkriyọṅ	ponkiriyoṅ	ponkriyọṅ	a sexual herbal medicine
33	gràgrà	gìràgìrà	gràgrà	gìràgìrà	gràgrà	being impatient
34	gìsh-gìsh	gìshì-gìshì	gìshì-gìshì	gìsh-gìsh	gìshgìsh	ticklish
35	skòròṅ	sikòròṅ	skòròṅ	sikòròṅ	skòròṅ	rascal
36	práprá	púrápúrá	práprá	púrápúrá	práprá	excellent
37	própró	púrópúró	própró	púrópúró	própró	excellent
38	jagbajańtìs	jagbajańtìsì	jagbajańtìs	jagbajańtìs	jagbajańtìs	rubbish
39	patakiláìs	patakiláìsì	patakiláìs	patakiláìs	patakiláìs	make important
40	kobaláìs	kobaláìsì	kobaláìs	kobaláìs	kobaláìs	implicate
41	brèkòyọṅ	bèrèkòyọṅ	brèkòyọṅ	bèrèkòyọṅ	brèkòyọṅ	bra
42	brèskòyọṅ	bèrèsikòyọṅ	brèskòyọṅ	bèrèsikòyọṅ	brèskòyọṅ	bra
43	intromímò	inturomímò	intromímò	inturomímò	intromímò	introduction
44	ìtráì	ìtíráì	ìtráì	ìtíráì	ìtráì	act of trying
45	láskábò	lásíkábò	láskábò	lásíkábò	láskábò	something enormous
46	skódà	síkódà	skódà	síkódà	skódà	smoothly
47	Sobońtóòṅ	Sobońtóòṅù	Sobońtóòṅ	Sobońtóòṅ	Sobońtóòṅ	brand name of an herb
48	fíryọyọ	fíríyọyọ	fíryọyọ	fíríyọyọ	fíryọyọ	grand
49	frijangban	fíríjangban	frijangban	fíríjangban	frijangban	unconstrained
50	fanimóròṅsì	fanimóròṅsì	fanimóròṅsì	fanimóròṅsì	fanimóròṅsì	attractive/appealing
51	gbeske	gbesike	gbeske	gbesike	gbeske	empty/bland
52	wòóskè	wòósískè	wòóskè	wòósískè	wòóskè	look (me) up
53	oníblòṅkí	oníbúlòṅkí	oníblòṅkí	oníbúlòṅkí	oníblòṅkí	block owner
54	gbaski-miski-yááskì	gbasiki-misiki-yáásikì	gbaski-miski-yááskì	gbasiki-misiki-yáásikì	gbaski-miski-yááskì	be a blast

## English forms

1	plug	púlòogì	plòogì	púlòòg	plòog
2	bread	búrédi	brédi	búrééd	bréèd
3	field	fíidi	fíidi	fíid	fíid
4	exam	èksáàmù	èksáàmù	èsáám	èksáám
5	slippers	sílípáàsì	slípáàsì	sílípáàs	slípáàs
6	globe	gílòòbù	glóòbù	gílòòb	glóòb
7	glass	gíláàsì	gláàsì	gíláàs	gláàs
8	six	sîsì	sîksì	sîs	sîks
9	belt	bèèti	bèti	bèèt	bèti
10	sex	sèèsi	sèksi	sèès	sèks
11	excess	èksèèsi	èksèèsi	èsèèes	èksèèes
12	bank	bánkì	bánkì	bánk	bánk
13	laptop	lápútòòpù	láp tòòpù	látòòp	láp tòòp
14	access	áksèèsi	áksèèsi	ásèèes	áksèèes
15	tools	túùsì	túùlsi	túùs	túùls
16	bolt	bóòtù	bóltù	bóòt	bólt
17	milk	mîkì	mîlkì	mîk	mîlk
18	hundred	ónndèèdi/ ónrèèdi	ónndrèèdi/óndrèèdi	ónndèèd/ ónrèèd	ónndrèèd/óndrèèd
19	whatsapp	wòtsaàpù	wòtsaàpù	wòsaàp	wòtsaàp
20	slate	sílèèti	slèèti	sílèèt	slèèt
21	coalpot	kolopóòti	kolpóòti	kolopóòt	kolpóòt
22	pampers	pámpáàsì	pámpáàsì	pámpáàs	pámpáàs
23	self	sèèfù	sèlfù	sèèf	sèlf
24	scarf	Síkáàfù	skáàfù	síkáàf	skáàf

(28) Cluster hypercorrection periphery of the lexicon (Grammars 5 and 6)

Input	Output		
	Grammar 5	Grammar 6	Gloss
<b>English forms</b>			
1 calculate	kakléèt	kakléèti	
2 jubilate	jublèèt	jublèèti	
3 formulate	fòmllèèt	fòmllèèti	
4 convenient	kòńfnènt	kòńfnènti	
5 cement	sméńt	sméńti	
6 paracetamol	parastamòò ~ prastamòò	parastamòò/prastamòò	
7 carburetor	kaplétò ~ kablétò ~ kabrétò ~ kaprétò	kaplétò/kablétò/kabrétò/kaprétò	
8 popular	póplà	póplà	
9 temporary	témprári	témprári	
10 university	unifèst	unifèsti	
11 supervisor	spáfáísò	spáfáísò	
12 president	présdèńt	présdèńti	
13 Superjet	Spájèèt	Spájèèti	name of a soccer club
14 privilege	priflèèj	priflèèji	
15 participate	patispèèt	patispèèti	
16 supermarket	spámakèèt	spámakèèti	
17 facility	fáslíti	fáslíti	
18 vicinity	fisníti	fisníti	
19 population	poplèşòn	poplèşòn	
20 ventilation	fèntléshòn	fèntléshòn	
22 intercontinental	intakòntnèńtál	intakòntnèńtà	
<b>Yoruba forms</b>			
1 Abèòkúta	Abèkta	Abèkta	name of a city
2 Abúlèègbá	Ablèègbá	Ablèègbá	name of a city
3 gbègìrì	gbègri	gbègri	bean soup
4 pakuru mò	pakru mò	pakru mò	pounce on
5 òfurufú	òfru.fú	òfru.fú	empty/sky
6 burúkú	brú.kú or burkú	brú.kú or burkú	bad
7 firì	frì	frì	sloppy
8 bó sí kòrò	bòskòrò	bòskòrò	second-hand cloth
9 bó sí kònà	bòskònà	bòskòn`à	second-hand cloth
10 bèsúkè	bèskè	bèskè	soybean cake
11 (Jáámíù) Jámí-sí-pátá	Jámíspátá	Jámíspátá	who defecates in his pants
12 àbùradà	àbradá	àbradá	umbrella
13 bùrùkùtù	brùkùtù or bùrkùtù	brùkùtù or bùrkùtù	a kind of beer
14 bírísopé	brísopé	brísopé	relating to bricklaying
15 kúrúkùsá	krúkùsá or kùrkùsá	krúkùsá or kùrkùsá	dandruff
16 ilòkùlò	ilòklò	ilòklò	bad invitation



17	dókítà	dóktà	dóktà	doctor
18	jambula	jambla	jambla	fumble
19	jó bùlà bùlà	jó blà blà	jó blà blà	burns wildly
20	àbùrò mi	àbro mi	àbro mi	my younger sibling
21	sùmèsùmè	smèsèmè	smèsèmè	slow
22	gìràgìrà	gràgrà	gràgrà	restless
23	sírísírí	srísrí	srísrí	ejecting bit by bit
24	àsikò	àskò	àskò	time
25	àbùrò	àbrò	àbrò	younger sibling
26	kùrò	kròò	kròò	move
27	kùrúkùsá	krúkùsá	krúkùsá	dandruff
28	tòm̀bùlà	tòm̀blà	tòm̀blà	local lantern
29	jaburata	jabrata	jabrata	plenty
30	kérésimesì	krésimesì	krésimesì	Christmas
31	bírìkótó	brìkótó	brìkótó	form alliance
32	fínrínfíntín	frínfíntín	frínfíntín	English
33	fùrùkòkò	frùkòkò	frùkòkò	fumble
34	àbùradà	àbradà	àbradà	umbrella
35	mòṣosáíkù	mòṣáíkù	mòṣáíkù	motorcycle
36	fíkánfíkán	fkánfkán	fkánfkán	smelling
37	fúkèfúkè	fkèfkè	fkèfkè	rise intermittently
38	fùjòfùjò	fjòfjò	fjòfjò	foam
39	mùsùlùmí	mùslùmí	mùslùmí	muslim
40	búlálà	blálà	blálà	whip
41	gbàrà	gbràà	gbràà	suddenly
42	tiketike	tketke	tketke	a kind of motorcycle
43	bírìkìlà	brìkìlà/brìklà	brìkìlà/brìklà	bricklayer
44	tí wọn	twọn	twọn	which they
45	káàkiri	káàkri	káàkri	everywhere
46	fúnra ara	frára ara	frára ara	oneself
47	bọ́ síta	bọ́sta	bọ́sta	go out
48	múmú láyà	múmláyà	múmláyà	important to
49	èmi nàà	èmnàà	èmnàà	me too
50	kí ló dé	klódé	klódé	what happened
51	sùgbọ́n	sgbọ́n	sgbọ́n	but
52	àtipé	àtpé	àtpé	and still
53	lá ti lé	látlè	látlè	in order to
54	ẹ́ kú rọ̀lẹ́	ẹ́ krọ̀lẹ́	ẹ́ krọ̀lẹ́	good evening
55	a sì tún	astùn	astùn	we can also
56	ìtẹ̀swájú	ìtẹ̀swájú	ìtẹ̀swájú	progress
57	Ááfúsá	Ááfsá	Ááfsá	a Muslim name
58	kùrúnà	krúnà	krúnà	dandruff
59	bírìsopé	brìsopé	brìsopé	bricklayer
60	gírìkì	grìkì	grìkì	authentic
61	gbọ̀mlélántà	gbọ̀mlélántà	gbọ̀mlélántà	cooperative loan
62	kálukú	káلكú	káلكú	everyone
63	ẹ̀ni tó	ẹ̀ntó	ẹ̀ntó	one who
64	fùlání	flání	flání	a Fulani person

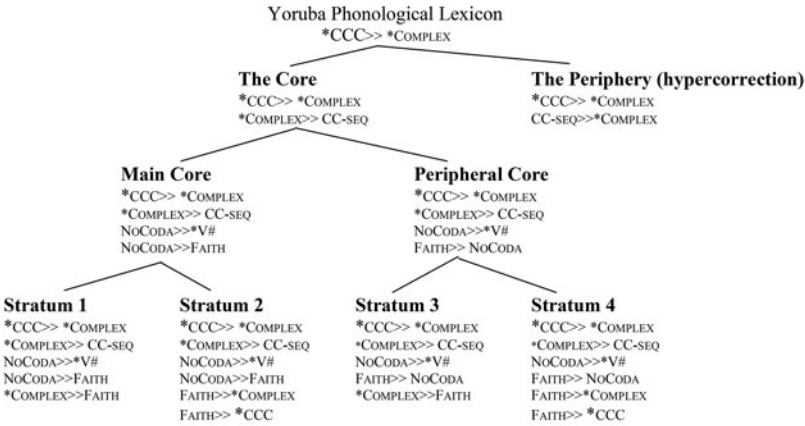
65 obìnrin	obrin	obrin	female
66 mósáláásí	mósláásí	mósláásí	mosque
67 dířẹ̀bà	drẹ̀bà	drẹ̀bà	driver
68 ẹ̀sámínà	ẹ̀sámnà	ẹ̀sámnà	examiner
69 kòmúsọ̀yọ̀n	kómsọ̀yọ̀n	kómsọ̀yọ̀n	bra
70 lébìrà	lébrà	lébrà	laborer

(29) Coda hypercorrection periphery of the lexicon (Grammar 7)

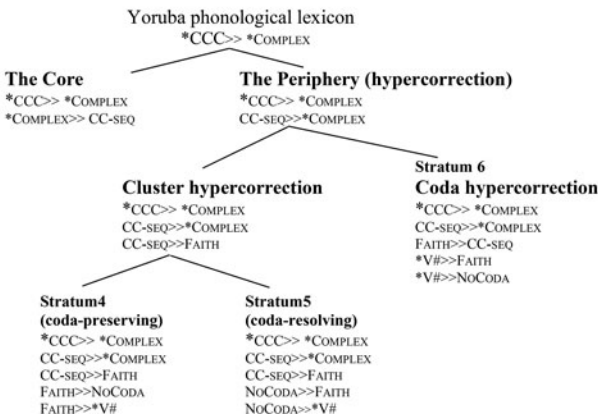
Input		Output	
		Grammar 7	Gloss
<b>English forms</b>			
1	NNPC / ɛnɛnpisi/	ɛnɛnpiis	
2	hospital	oʃpiit	
3	area	erial	
<b>Yoruba forms</b>			
1	Şàgámù	Şàgáàm	name of a city
2	oşòdì	oşòòd	an area in Lagos
3	sáàşì	sáàs	search
4	ìgbádùn	ìgbaàd	enjoyment
5	gba	gbààs	take
6	jedijedi	jedijeed	piles
7	ó yá	ó yáás	it's time
8	wo bí	wo bíis	look here
9	wá	wáás	come
10	bátìrì	bátir	battery
11	miliiki	miliik	milk
12	túláàsì	túláàs	compulsory
13	ó pò	ó pòr [ó kpòr]	it's plenty
14	şòòkì	şòòk	chalk
15	bélîtù	bélît	belt
16	méjéèjì	méjèej	the two of them
17	şóòşì	şóoş	church
18	bèè ni	bèèn	yes
19	máà mi	máàm	my mum
20	báà mi	báàm	my dad
21	bòòdá mi	bòòdám	my brother
22	ẹ̀ káàbò	ẹ̀kààb	welcome
23	kí ni	kín	what is it?
24	ìyá kókó	ìyá kókós	woman who sell koko
25	mọ̀kálîkì	mọ̀kálîk	mechanic

APPENDIX C: GRAMMAR LATTICE FOR A PARTIAL-ORDER ANALYSIS

(30) a. Yoruba GRAMMAR lattice (Core)



b. Yoruba GRAMMAR lattice (Periphery)



**APPENDIX D: GRAMMARS 2–7 IN A LEXICALLY INDEXED PARTIAL-ORDER CO-PHONOLOGY**

(31) a. Grammar 2: NoCODA >> \*V#... >> FAITH >> \*CCC >> \*COMPLEX >> CC-SEQ

	NoCODA	*V#...	FAITH	*CCC	*COMPLEX	CC-SEQ
I. /brɛd/s <sub>1-7</sub>	(a) [bú.rɛ̄.di]	*	**DEP!			*
	(b) [brɛ̄.di]	*	*DEP		*	
	(c) [brɛ̄.ɛ̄d]	*!			*	
II. /onibàtə/s <sub>1-4</sub>	(a) [o.ní.bà.tà]	*				*
	(b) [on.í.bà.tà]	*!	*			*
	(c) [o.nbà.tà]	*	**MAX!		*	
	(d) [o.ní.bàt]	*!	**MAX			*
	(e) [o.nbtà]	*	****MAX	*!	**	

b. Grammar 3: \*CCC >> \*COMPLEX >> FAITH >> \*V#, CC-SEQ, NoCODA

	*CCC	*COMPLEX	FAITH	*V#	CC-SEQ	NoCODA
/plʌg/s <sub>1-7</sub>	(a) [pú.lò.òg]		*DEP		*	*
	(b) [plò.òg]		*!			*
	(c) [plò.ò.gì]		*!	*DEP	*	
	(d) [pú.lò.ò.gì]			**DEP!	*	*

## c. Grammar 4: FAITH &gt;&gt; \*CCC &gt;&gt; NoCODA &gt;&gt; \*V#... &gt;&gt; CC-SEQ &gt;&gt; \*COMPLEX

		FAITH	*CCC	NoCODA	*V#...	CC-SEQ	*COMPLEX
I. /brɛd/s <sub>1-7</sub>	(a) [bú.rɛ̀.dì]	**DEP!			*	*	
	(b) [brɛ̀.dì]	*DEP!			*		*!
	⊙ (c) [brɛ̀.ɛ̀d]			*			*
II. /oníbàtá/s <sub>1-4</sub>	⊙ (a) [o.ní.bà.tà]				*	*	
	(b) [on.í.bà.tà]			*!	*	*	
	(c) [o.nbà.tà]	*MAX!			*		*!
	(d) [o.ní.bàt]	*MAX!		*!		*	
	(e) [o.nbtà]	**MAX	*!		*		**

## d. Grammar 5: CC-SEQ &gt;&gt; FAITH &gt;&gt; \*V#... &gt;&gt; NoCODA &gt;&gt; ...\*CCC &gt;&gt; \*COMPLEX

		CC-SEQ	FAITH	*V#...	NoCODA	...*CCC	*COMPLEX
I. /dzu:biłerit/s <sub>1-7</sub>	(a) [ju.bi.lé.èt]	*!			*		
	(b) [ju.bi.léè.ti]	*!	*DEP	*			
	(c) [ju.bléè.ti]		*MAX, *DEP!	*			*
	⊙ (d) [ju.blé.èt]		*MAX		*		*
	(e) [ju.bi.lé]	*!	*MAX	*			
II /gbègiri/s <sub>1-6</sub>	(a) [gbè.gì.ri]	*!		*			
	⊙ (b) [gbè.gri]		*MAX	*			*
	(c) [gbè.gì.ris]	*!	*DEP		*		
	(d) [gbè.gris]		*MAX, *DEP!		*		*

e. Grammar 6: NoCODA >> ...CC-SEQ, \*V#...>> FAITH...>> ...\*CCC >> \*COMPLEX

		NoCODA	...CC-SEQ	*V#...	FAITH...	*CCC	*COMPLEX
I. /dʒu:bɪlɛt/s1-7	(a) [ju.bi.lé.èt]	*!	*				
	(b) [ju.bi.lé.è.ti]		*!	*	*DEP		
	☞ (c) [ju.blé.è.ti]			*	*MAX, *DEP		*
	(d) [ju.blé.èt]	*!			*MAX		*
	(e) [ju.bi.lé]		*!	*	*MAX		
II. /gbɛ̀giri/s1-6	(a) [gbɛ̀.gi.ri]		*!	*			
	☞ (b) [gbɛ̀.gri]			*	*MAX		*
	(c) [gbɛ̀.gi.ris]	*!	*		*DEP		
	(d) [gbɛ̀.gris]	*!			*MAX, *DEP		*

f. Grammar 7: \*V#>> NoCODA (...FAITH, CC-SEQ)>> FAITH>> CC-SEQ>>...\*CCC >> \*COMPLEX

		*V#	NoCODA (...FAITH, CC-SEQ)	FAITH	CC-SEQ	*CCC	*COMPLEX
/bɛ̀ɛ̀ ni/s1-4,7	(a) [bɛ̀.ɛ̀.ni]	*!			*		
	☞ (b) [bɛ̀.ɛ̀n]		*	*MAX	*		
	(c) [bɛ̀.ɛ̀]	*!		**MAX	*		
	(d) [brɛ̀.ɛ̀n]		*	*MAX, *DEP!			*
	(e) [brɛ̀.ɛ̀.ni]	*!		*DEP			*