

# The Role of Underlying Representations in L2 Brazilian English

Andrew Nevins (Harvard), David Braun (MIT)

## 1. Overview

---

In this paper we examine two phenomena in the phonology of the English spoken by native speakers of Brazilian Portuguese (henceforth, Brazilian Portuguese English: BPE). The first is the phenomenon of “spurious affrication” before [u], a phenomenon in which English sequences of [tu] are rendered [tʃu] in BPE as well as in English loanwords adapted into Brazilian Portuguese. In discussing this phenomenon we provide additional background and exemplification of affrication in BP and BPE. The second phenomenon we discuss is “rhotic hypercorrection”, a phenomenon in which English word-initial [h] is rendered in BPE as [r]. Spurious affrication and rhotic hypercorrection present a problem for models of loanword phonology such as LaCharité and Paradis (2005), which propose that speakers map the L2 input onto the closest phonological analogue in their L1, since BP contains the sequences [tu] and initial [h] in its native phonology. These phenomena are especially interesting in light of the fact that both phonetic approximation and the orthography of English would militate against spurious affrication and rhotic hypercorrection. We propose an explanation in terms of the underlying representations (URs) that BPE speakers adopt, thereby highlighting a crucial role for URs in L2 and loanword phonology.

The data discussed here come from a variety of sources, and reflect both systematic and sporadic occurrences, observed both in Boston’s large Brazilian immigrant community and in various cities in Brazil. The phenomena described here may occur both in formal and casual speech registers. The phenomena of spurious affrication and rhotic hypercorrection were noticed independently by both authors and the salience of these observations, heretofore unexplored in the literature, provided the inspiration for the present paper. We have noticed no particular demographic variables that correlate with the occurrence of these phenomena. Due to the composition of the Boston Brazilian community, most of the observations reported here were produced by speakers originally from the state of Minas Gerais.

## 2. Spurious Affrication

---

In this section we report on the existence of BPE productions in which the sequence *tu* in English surfaces with the affricate [tʃu] instead. To understand why this puzzling and BPE-specific L2 phenomenon occurs, it is necessary to introduce first the distribution of affrication within the native phonology.

## 2.1. Background on BP Affrication

---

BP has the seven oral vowels in (1), as well as nasalized counterparts of the cardinal five.

- (1) i u  
e o  
ɛ ɔ  
a

The rule of affrication in BP is much like that of other languages:

- (2) Affrication: /t,d/ → [tʃ,dʒ] / - [-cons, +high, -back]

The effects of affrication can be seen in the following examples. In the transcriptions that follow, we include the effects of vowel reduction, raising a final unstressed /o/ to [u] and /e/ to [i]. Note that vowel reduction feeds affrication.

- (3) a. *ticket* [tʃi.ke.tʃi] ‘ticket’  
b. *tirar* [tʃi.rax] ‘to take out’  
c. *mente* [mẽ.tʃi] ‘mind’  
d. *sede* [se.dʒi] ‘thirst’  
e. *Diogo* [dʒi.o.gu] (Proper name)  
f. *mastigar* [mas.tʃi.gax] ‘to chew’  
g. *cestinha* [ses.tʃĩ.na] ‘basket (dim.)’

Affrication in BP is described in Cagliari (1997). For discussion of the phonology and phonetics of affrication more generally, see Calabrese (2005), Hall and Hamann (2006), and Kim (2001). Crucially, *affrication in BP does not apply before the vowel [u]*.

- (4) a. *tucano* [tu.kẽ.nu] ‘toucan’  
b. *turma* [tuɣ.ma] ‘group’  
c. *costurar* [kos.tu.rax] ‘to sew’  
d. *mistura* [mis.tu.ra] ‘mixture’  
e. *gato* [ga.tu] ‘cat’

## 2.2. Spurious Affrication in BPE

---

The curious phenomenon of BPE spurious affrication is the following: BPE speakers produce English words with the sequence [tu] as [tʃu], *even though affrication does not apply before [u] in BP*:

- (5) Spurious affrication of English [tu]-sequences:  
a. *to, two* [tʃu]  
b. *U2* [ju.tʃu] or [iw.tʃu]  
c. *student* [stʃu.dənt] or [is.tʃu.dənt]  
d. *stew* [stʃu] or [is.tʃu]  
e. *stupid* [stʃu.pɪd] or [is.tʃu.pɪd]

- f. *during* [dʒu.rɪŋ]

Our proposal is that spurious affrication results from an underlying representation in which these sequences are not, in fact, represented as /tʊ/. It is a fact that the English [u] is far more fronted than the BP [u], particularly after coronal consonants. It is probably more appropriate to transcribe the English back rounded vowel as [tʊ] (or [tʰ]). How does a BPE speaker deal with an incoming token of English [tʊ] (or [tʰ])? Contra the predictions of LaCharité and Paradis (2005), they are not simply matching it with the BP [u]. This is a pervasive loanword phenomenon. Any radio announcer or schoolkid in Brazil will pronounce the name of the band U2 as [ju.tʃu], with an affricate. BPE speakers are attempting to approximate the phonetic realization of [tʊ] (or [tʰ]). However, our proposal is that they must do so *using the resources available to them in their native language*. Since BP does not contain [tʊ] (or [tʰ]), those are not an option for the underlying form of *two*. We propose that BPE speakers approximate the fronted quality of English [tʊ] (or [tʰ]) by setting up an underlying representation with a non-nuclear [i].<sup>1</sup> The underlying forms of the words in (5) are thus the following:

- (6) BPE speakers' URs for English [tʊ]-sequences:
- a. *to, two, too* /t̞iu/
  - b. *U2* /i̞u.t̞iu/
  - c. *student* /st̞iu.dənt/
  - d. *stew* /st̞iu/
  - e. *stupid* /st̞iu.pɪd/
  - f. *during* /d̞iu.rɪŋ/

In fact, the use of /i̞u/ to approximate post-coronal allophones of English [u] in the underlying representations of BPE is not unique to the stops. BPE speakers adopt /i̞u/ for many post-coronal occurrences of English /u/. It is only with the stops that the non-nuclear high front vocoid /i̞u/ triggers affrication. With other preceding coronals, the /i̞u/ remains, and many speakers enact a process of "fusion" whereby the [−back, +high] non-nuclear segment and the [+round, +high] nucleus fuse into a front rounded vowel:

- (7)
- a. *new* /n̞iu/ → [n̞ü]
  - b. *soon* /s̞iun/ → [s̞ün]
  - c. *noon* /n̞iun/ → [n̞ün]

The fact that the sequence /i̞u/ triggers affrication in the case of a preceding stop but triggers nuclear fusion in the case of a preceding nasal or fricative suggests a context-sensitive resolution of the same underlying sequence in different ways, suggesting that loanwords are being adopted according to an active model of "analysis-by-synthesis", as proposed by Calabrese (2006).

<sup>1</sup>Some of our consultants, when asked why they produce spurious affrication, tell us that they hear an *i-zinho* 'a little i' in English words like *two*.

A word is in order here about other possible explanations for spurious affrication. One might think that spurious affrication in words like *two* (English [t<sup>h</sup>u]) represents an attempt to approximate the aspiration of the voiceless stop by using the resources of the native language, i.e. the turbulence of an affricate. While interesting, this possibility stops short when it comes to *st* words like *student*, which are not aspirated in English. It also could not be extended to the cases of affrication with voiced coronal stops, such as *during*.<sup>2</sup> Finally, we dismiss the possibility that representations such as /tʃ/ come from hearing dialects of English that allow [tʃu] in stressed syllables, such as British English, as our consultants' exposure is almost exclusively to American English, which disallows [tʃu] in stressed syllables (McCarthy and Taub, 1992), as American English is the dialect taught in Brazilian schools, and finally, because even British English does not have [tʃu] for all of the words above.

We propose that BPE speakers apply rules of their native BP to underlying forms they have set up. The application of Affrication (2) to the forms in (7) will thus yield spurious affrication.

The interest of this phenomenon is the fact that, as Peperkamp and Dupoux (2003) suggest, L2 and loanword phonology does involve an attempt to approximate the phonetic form of the donor language. However, in the case at hand, this approximation is *done through the phonology, and achieved by setting up an underlying form which contains the phonetic approximation*. Once this UR is set up, it is subject to the automatic rule of affrication just like any other underlying sequence of coronal plus high front vowel.

### 3. Rhotic Hypercorrection

---

In this section we report on the existence of BPE productions in which English *h*-initial words surface with an initial *r*. To understand why this puzzling and BPE-specific L2 phenomenon occurs, it is necessary to introduce first the distribution of rhotics within the native phonology.

#### 3.1. Background on BP rhotics

---

We begin by reviewing the phonology of rhotics in BP. BP has three basic rhotics: [r, ʁ, x], whose surface distribution is the following:

- (8) [ʁ] occurs syllable-initially when not postvocalic:<sup>3</sup>
- a. *ra*bo [ʁabu] 'tail'
  - b. *re*i [ʁej] 'king'
  - c. *ro*quenrou [ʁõ.kẽ.frow] 'rock and roll'
  - d. *hon*ra [õ.ʁa] 'honor'

---

<sup>2</sup>Affrication with the voiced stop is rarer in our observations, but this may reflect our limitations as observers, since the fricative portion of [dʒ] is less salient than that of [tʃ].

<sup>3</sup>We assume that words like *honra* have an underlying nasal coda consonant that is deleted, following Mattoso Câmara Jr. (1970) and Wetzels (1997). This coda consonant conditions the allophony of the following postconsonantal [ʁ].

- e. *israel* [iz.ʃia.ɛw] ‘Israel’
  - f. *dahruj* [da.ʃiuʒ] (Proper name)
- (9) [x] occurs in the coda:<sup>4</sup>
- a. *mar* [max] ‘ocean’
  - b. *carne* [cax.ni] ‘meat’
  - c. *circo* [six.ku] ‘circus’
- (10) [r] occurs in complex onsets:
- a. *prato* [pra.tu] ‘plate’
  - b. *abre* [a.bri] ‘open!’
  - c. *freio* [frej.ju] ‘brake’
- (11) [r] and [x] contrast intervocalically:<sup>5</sup>
- a. *carro* [ca.yu] ‘car’
  - b. *caro* [ca.ru] ‘dear’
  - c. *barra* [ba.ɣa] ‘bar’
  - d. *barato* [ba.ra.tu] ‘cheap’

The literature on whether the fricatives and the rhotic are simply distinct phonemes or are allophones in (near-)complementary distribution is vast. We adopt the view, following Mattoso Câmara Jr. (1953), Lopez (1979), Oliveira (1997), Mateus and d’Andrade (2000) and Abaurre and Sandalo (2003) that all of these surface allophones reflect a single underlying phoneme, which we posit is [r].

Many lines of evidence point towards this conclusion. The first comes from affixation and sandhi phenomena, which demonstrate that a coda [x] can become an [r] when followed by a vowel

- (12) a. *por* [pox] ‘through’  
b. *por cima* [pox.sĩ.ma] ‘through above’  
c. *por aqui* [po.ra.ki] ‘through here’
- (13) a. *flor* [flox] ‘flower’  
b. *flores* [flores] ‘flowers’

The converse occurs when an intervocalic [r] becomes [ʃ] when truncation occurs:

- (14) a. *direto* [dʒiɾetu] ‘straight’  
b. *reto* [hetu] ‘straight, straight on’

We propose that the underlying rhotic is thus the tap [ɾ], which undergoes the following rules:

---

<sup>4</sup>There is a wide range of sociophonetic and variationist work on the realization of coda rhotics in BP, which may be more velar, uvular, or glottal, depending on a variety of demographic and geographic factors; see, among others, Dinah Callou and Yonne Leite and João Moraes (2002). As our focus here is on word-initial rhotics, this variation lies outside the scope of the current problem.

<sup>5</sup>Fricatives are subject to intervocalic voicing in BP, hence the allophone [x] undergoes a further change to [ɣ]. Regressive voicing in consonantal sequences also yields [ɣ] in words like *turma* [tuɣ.ma].

- (15) Onset debuccalization:  $r \rightarrow \text{fi}$  in  $\{\#, C\}.$ –  
 (16) Coda spirantization:  $r \rightarrow x$  in  $.\{\#, C\}$

Finally, we analyze intervocalic  $[x]$  as the result of a heterosyllabic geminate (see, e.g. Harris (1983, 2002) for Spanish. Note that each half of the geminate will undergo one of the rules in (15) and (16).

### 3.2. BPE rhotic hypercorrection

---

We turn to the phenomenon of central interest in this section: the fact that BPE speakers occasionally produce tokens such as the following:

- (17) Sporadic hypercorrection of English *h*-initial words:  
 a. *home*  $[\text{rom}] \sim [\text{hom}]$   
 b. *hug*  $[\text{r}\Lambda\text{g}] \sim [\text{h}\Lambda\text{g}]$   
 c. *hunger*  $[\text{r}\tilde{\Lambda}\text{g}\text{ər}] \sim [\text{h}\tilde{\Lambda}\text{g}\text{ər}]$   
 d. *hammock*  $[\text{r}\text{æm}\text{ək}] \sim [\text{h}\text{æm}\text{ək}]$

The data on the left-hand side of the squiggle above are quite surprising at first blush. Why would a speaker who has heard the English word *home* pronounced as  $[\text{hom}]$  suddenly make the decision to pronounce it with an initial  $[\text{r}]$ ? The descriptive answer is, this is hypercorrection. But what is the mechanism for hypercorrection, and how do the data in (17) fit within a broader theory of L2 and loanword adaptation?

We would like to make the proposal that *BPE speakers analyze incoming English words in terms of their native BP phonology, and set up underlying representations based on their native phonology*. In this respect our proposal resembles that of Ito et al. (2006), who propose that loanword adaptation involves the application of one’s native inventory to the incoming data. In particular, we propose that the words in (17) have been set up with the URs in (18)

- (18) BPE speakers’ URs for English words:  
 a. *home*  $/\text{rom}/$   
 b. *hug*  $/\text{r}\Lambda\text{g}/$   
 c. *hunger*  $/\text{r}\tilde{\Lambda}\text{g}\text{ər}/$   
 d. *hammock*  $/\text{r}\text{æm}\text{ək}/$

Given the URs in (18), what accounts for the variable data in (17)? Our proposal is that speakers variably apply Onset Debuccalization (15) when speaking English. *The hypercorrected forms thus reflect the naked URs that these speakers have adopted.*

In fact, our proposal here resembles to some extent the assertion of Peperkamp and Dupoux (2003) that in L2 and loanword approximation, speakers attempt to match up the phonetics of the donor language as closely as possible. The difference is that the route to setting up this match is through the UR. BPE speakers who hear an English word-initial  $[\text{h}]$  indeed attempt to produce a surface  $[\text{h}]$  in their output grammars, but do so by setting up an underlying  $/\text{r}/$

and letting the regular rules of BP underlying-to-surface mapping do the work of achieving the target output. Occasionally, however, these rules of UR-to-SR mapping fail, revealing the UR that BPE speakers have set up in their attempt to achieve the correct surface approximation of the English words.

A question arises at this point as to why the allophonic rule of Initial Debuccalization sometimes “fails” for the URs set up in (18). Why should the rule of Initial Debuccalization apply variably to these representations, when it applies categorically to native BP words?

The discussion thus far has only focus on English words that are surface *h*-initial, and we have not put forth yet a proposal as to the underlying representations of English words that are, in fact, surface *r*-initial, such as *radar*. Given the fact that BP has no possibility of a surface initial tap, what could the underlying representation be? Our proposal is that words such as *radar* are set up with an underlying tap, but are furthermore marked as exceptions to Initial Debuccalization. That is to say, on an item-by-item basis, these words must be diacritically annotated as exempt from the otherwise operative rule. Two representative grammatical models of exception-marking, are Zonneveld (1978) in a generative phonology tradition and Pater (2006) in terms of lexically-specific faithfulness constraints in OT. For ease of exposition, we adopt the former here:

- (19) a. *radar* /rejd̥ar/ [−InitialDebucc]  
 b. *respect* /rispekt/ [−InitialDebucc]  
 c. *Redford* /rɛdfɔrd/ [−InitialDebucc]

There is, to a limited extent, already precedent for exception-marking to allophonic rules in the native phonology of BP. One example comes from truncation of words with intervocalic rhotics. The cocktail known as a *caipiroska*, made with crushed limes, has undergone a number of adaptations with different fruits, such as *morangoroska*, made with crushed strawberries. Much like the development of the English morpheme *-tini* (originally from *martini*, but now found in such coinages as *appletini* and *chocolatetini*), a truncated form *roska* has emerged as a catch-all term for cocktails such as *caipiroska*, *morangoroska*, *man-garoska*. Interestingly, this word is pronounced as [rɔska], with an initial tap. Given that this word stands as an exception to Initial Debuccalization, it must have the UR in (20-c):

- (20) a. *caipiroska* [kajpirɔska]  
 b. *roska* [rɔska], \*[fioska]  
 c. /roska/, [−InitialDebucc]

Exceptionally-faithful underlying forms resistant to allophonic rules can also be found with the otherwise extremely general and productive rule of Liquid Fronting in the plural, shown in (21):

- (21) a. *jornal* [ʒornaw, ʒornaʒs] ‘newspaper, sg. & pl.’  
 b. *radical* [fiadʒikaw, fiadʒikaʒs] ‘radical, sg. & pl.’  
 c. *hotel* [otew, oteʒs] ‘hotel, sg. & pl.’

- d. *pastel* [pastew, pastejs] ‘pastry, sg. & pl.’
- e. *caracol* [karakow, karakois] ‘coil, sg. & pl.’
- f. *sol* [karakow, sojs] ‘sun, sg. & pl.’
- g. *anzol* [karakow, êzojs] ‘hook, sg. & pl.’
- h. *lençol* [lêsoiw, lêsojs] ‘layer, sg. & pl.’

Liquid Fronting may be formulated as follows:

- (22) [+liquid] → [−consonantal, +high, −back] / \_+s#

Nonetheless, at least two nouns in BP (23) must be marked as [−LiquidFronting]:

- (23) a. *gol* [gow, gowz] ‘goal, sg. & pl.’  
b. *Skol* [skow, skowz] ‘Skol (brand of beer), sg. & pl.’

Thus, the existence of exceptional non-undergoing has ample precedent in the phonology of BP. Returning to the representation of *h*-initial vs. *r*-initial words in English, we may now contrast the two divergent underlying representatins of *home* and *rome*, for example:

- (24) BPE speakers’ URs for English words:  
a. *home* /rom/  
b. *Rome* /rom/ [−InitialDebucc]

As noted by Zonneveld (1978) and many others, the existence of exception-marking diacritics may be somewhat grammatically fragile. We propose that the existence of minimal pairs such as (24-a)-[b] that differ only in the presence of a rule-non-undergoing diacritic are what lead to the occasional inhibition of Initial Debuccalization that gives rises to Rhotic Hypercorrection. Naturally, the occasional omission (or failure-to-access) such diacritics also may give rise to occasional BPE productions of words like *radar* with an initial [h].

Most crucially for the existence of Rhotic Hypercorrection, however, is the occurrence of a UR that is set up in accordance with the native phonotactics of BP in mind and subject to the normal rule of Initial Debuccalization, which will normally yield an output identical to the intended English word. That is, with an “incorrect” (or, more neutrally, *divergent*) UR, BPE speakers usually arrive at the correct target output for English *h*-initial words. Sporadic suppression of the allophonic rule producing this convergent output reveals the otherwise hidden divergent UR.

#### 4. Conclusion

---

Both spurious affrication and rhotic hypercorrection are puzzling phenomena from the perspective of either matching the surface phonetics or choosing the most identical underlying match for the donor language’s correspondent. We have proposed a new model of L2 and loanword phonology here, one in which the speakers indeed attempt to match the surface forms of the donor language, e.g. English [tʰ] (or [tʰ]) and [hom], but must do so using the URs of their



native phonology, viz., /tʲiu/ and /rom/. The crucial factor that yields spurious affrication and rhotic hypercorrection is the (sometimes abstract) form that words are stored in long-term memory during an early stage of contact with the donor language. When the UR-to-SR allophonic rules of the native language are left to take over, the surface result may end up different from the input.

## References

---

- Abaurre, Bernadete, and Filomena Sandalo. 2003. Os róticos revisitados. In *Teoria Lingüística: fonologia e outros temas*, 144–180. João Pessoa: Editora Universitária da UFPB.
- Cagliari, Luiz Carlos. 1997. *Fonologia do Português: Análise pela geometria de traços*. Campinas: Edição do Autor.
- Calabrese, Andrea. 2005. *Markedness and Economy within a Derivational Model of Phonology*. Mouton de Gruyter.
- Calabrese, Andrea. 2006. An Analysis-by-Synthesis Model of Loanword Phonology. Paper presented at Going Romance, Amsterdam.
- Dinah Callou and Yonne Leite and João Moraes. 2002. Processo(s) de Enfraquecimento Consonantal no Português do Brasil. In *Gramática do Português Falado VIII*, 537–555. Campinas: Editora da UNICAMP.
- Hall, Tracy, and Silke Hamann. 2006. Towards a typology of stop assibilation. *Linguistics* 44.6:1195–1236.
- Harris, James W. 1983. *Syllable Structure and Stress in Spanish*. MIT Press.
- Harris, James W. 2002. Flaps, Trills, and Syllable Structure in Spanish. In *Phonological Answers (and their Corresponding Questions)*, MITWPL 42, 81–108. Cambridge, MA: MITWPL.
- Ito, Chiyuki, Yoonjung Kang, and Michael Kenstowicz. 2006. The Adaptation of Japanese Loanwords into Korean. In *Studies in Loanword Phonology*, MITWPL 52, 65–104. Cambridge, MA: MITWPL.
- Kim, Hyunsoon. 2001. A phonetically based account of phonological stop assibilation. *Phonology* 18:81–108.
- LaCharité, Darlene, and Carole Paradis. 2005. Category Preservation versus Phonetic Approximation in Loanword Adaptation. *Linguistic Inquiry* 36.2:223–258.
- Lopez, Barbara. 1979. The Sound Pattern of Brazilian Portuguese (Cariocan Dialect). Doctoral Dissertation, UCLA.

- Mateus, Helena, and Ernesto d'Andrade. 2000. *The Phonology of Portuguese*. Oxford University Press.
- Mattoso Câmara Jr., Joaquim. 1953. *Para o estudo da fonêmica portuguesa*. Rio de Janeiro: Organizações Simões.
- Mattoso Câmara Jr., Joaquim. 1970. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.
- McCarthy, John, and Alison Taub. 1992. Review of Paradis & Prunet (eds.), *The Special Status of Coronals*. *Phonology* 9:363–370.
- Oliveira, Valéria. 1997. Um reestudo da vibrante: análise variacionista e fonológica. Doctoral Dissertation, Pontifícia Universidade Católica do Rio Grande do Sul.
- Pater, Joe. 2006. The Locus of Exceptionality: Morpheme-Specific Phonology as Constraint Indexation. ROA-866.
- Peperkamp, Sharon, and Emmanuel Dupoux. 2003. Reinterpreting Loanword Adaptations: The Role of Perception. In *Proceedings of the 15th International Conference of Phonetic Sciences*, 367–370. .
- Wetzels, Leo. 1997. The Lexical Representation of Nasality in Brazilian Portuguese. *Probus* 9.2:203–232.
- Zonneveld, Wim. 1978. *A Formal Theory of Exceptions in Phonology*. Peter de Ridder Press, Lisse.