

# Vowel insertion after coda plosives in the dialect of Paraíba (PB): reflections on the establishment of syllabic and social boundaries

Felipe Santos dos Reis\*

Rubens Marques de Lucena\*\*

**Resumo:** O presente estudo buscou verificar como falantes do português brasileiro da cidade de Campina Grande (PB) produzem oclusivas em coda nas posições medial e final de palavras, considerando as diferentes possibilidades de lidar com consoantes não licenciadas em posição de coda. Para tanto, 960 ocorrências das oclusivas /p, t, k, b, d, g/ em coda foram gravadas em áudio e submetidas à análise estatística. Os resultados indicam que a inserção vocálica é influenciada por consoantes perdidas e pertencimento dos vocábulos ao acervo lexical, e dois fatores de natureza extralinguística, quais sejam, gêneros e níveis de proficiência em L2.

**Palavras-chave:** Consoantes perdidas em coda. Inserção vocálica. Falantes campinenses de PB. Fatores linguísticos e extralinguísticos.

**Abstract:** The present study aimed to verify how Brazilian Portuguese (BP) speakers born in the city of Campina Grande (PB) produce coda plosives in word-medial and -final positions, considering the possibilities of dealing with unlicensed consonants in coda position. Thus, 960 tokens of syllables closed by /p, t, k, b, d, g/ were recorded in audio and submitted for multivariate analysis. The results indicate that the application of the vowel insertion rule is influenced by the stray consonants and the words' etymology, and two extralinguistic ones, that is, speakers' gender and their L2 proficiency levels.

**Keywords:** Stray consonants in coda. Vowel insertion. Native BP speakers. Linguistic and extralinguistic factors.

**Resumen:** El presente estudio buscó determinar cómo los hablantes de portugués brasileño de la ciudad de Campina Grande (PB) producen las oclusivas en coda en posición media o final de palabras, en vista de las posibilidades para hacer frente a las consonantes que no pueden ocupar tal posición. Así, 960 ocurrencias de las oclusivas /p, t, k, b, d, g/ en coda fueron grabadas en audio y sometidas al análisis estadístico. Los resultados indican que la inserción vocálica es influenciada por los elementos extraviados y la etimología de las palabras, además de los géneros y niveles de competencia en L2.

\* Doutor em Linguística pelo Programa de Pós-Graduação em Linguística da Universidade Federal da Paraíba. <http://orcid.org/0000-0002-7073-2448>

\*\* Professor Associado da Universidade Federal da Paraíba, onde atua no Departamento de Letras Estrangeiras Modernas (DLEM) e no Programa de Pós-Graduação em Linguística (PROLING). <http://orcid.org/0000-0003-0799-299X>



**Palabras clave:** Consonantes perdidas en coda. Inserción vocálica. Hablantes campinenses de PB. Factores lingüísticos y extralingüísticos.

## Introduction

Based on the definition of epenthesis as an “addition of a phonetic segment in word-medial position” (XAVIER; MATEUS, 1990, p. 146 apud MASSINI-CAGLIARI, 2015, p. 283), it is possible to observe several types of structures which are likely to trigger the emergence of an epenthetic vowel (which usually surfaces as [i] in the context of pretonic and tonic syllables or [ɪ] in posttonic syllables), due to syllable structure violations. Brazilian Portuguese (henceforth, BP), for instance, allows a limited set of segments in codas: they must be [-voc, +son] or [-son, +cont, +cor] to fill such a position in our language (ALVES, 2004). When illegal consonants occur in coda position, they are referred to as “stray consonants” (C’), which means that these segments are not associated to any syllabic node, because they are not allowed in syllable-final position (COLLISCHONN, 2004).

Unlike English, stops are not licensed to occur syllable-finally in BP. Therefore, when an unsyllabified segment occupies the postvocalic slot, some variable processes can take place: (1) the word may *undergo epenthesis and resyllabification* (ad.vo.ga.do → a.d[i].vo.ga.do); (2) the underlying plosive consonant may be deleted by Stray Erasure (ha.bi.tat → ha.bi.ta[∅]); or (3) the stray consonant may surface in the original position (op.tar → op[∅].tar) by means of the Coda Condition<sup>1</sup> Weakening (CCW).

Epenthesis, as a syllable-based process that saves stray consonants, is a typical phonological phenomenon observed in a number of languages, including BP, where it is part of the syllabification mechanism and is usually triggered when syllable well-formedness conditions are violated. In fact, Barbosa (2004, p. 227) claims that “[...] BP exhibits vowel epenthesis, breaking most two-obstruent sequences.” Therefore, words like “*psicologia*” (“psychology”) and “*advogado*” (“lawyer”) are likely to be produced with an additional high front vowel (either [i] or [ɪ]) between their inter- and intrasyllabic obstruents. Yet, there is another variable process in some Brazilian dialects, referred to

---

<sup>1</sup>Coda condition is also referred to as “coda filter”.

as “Coda Condition Weakening” (CCW), acting on obstruents in syllable-final position: words like “*fac.ção*” (“faction”), “*pac.to*” (“pact”), “*op.ção*” (“option”) e “*rép.til*” (“reptile”), for example, may be heard without the addition of phonetic substance, since the labial and dorsal plosives can be incorporated into the coda by the CCW. The fact that coda obstruents may surface in some BP dialects is explained by Bisol (1999, p. 731) as a “[...] remnant of an old grammar.”

Considering the dearth of investigations focusing on northeastern varieties of BP (CARDOSO, 2005; LUCENA, 2012), this sociolinguistic study aims to examine a variable pattern in the dialect of Paraíba (PB) by focusing on the production of coda plosives in word-medial and -final position by Brazilian L2 speakers from the city of Campina Grande (PB). Regarding the choice of a speech used in the Northeast region of Brazil, in addition to the aforementioned scarcity of studies, it is also important to point out the fact that since its emergence as a research paradigm in the late 1960s and early 1970s, the variationist sociolinguistic approach has demonstrated a special interest in the scientific study of stigmatized, non-standard varieties as a way of refuting linguistic stereotypes (cf. BAYLEY, 2005).

Our interest in coda plosives is motivated not only by the limited amount of research focusing on the dialects of the Northeast region in general, but also by the need to better understand how these structures are produced: it is still necessary to verify if epenthesis is a categorical process, as suggested by John and Cardoso (2017, p. 181), even though the authors admit that “[...] some variation has been observed in the phonetic realization of epenthetic [i] in BP, such that the vowel is phonetically gradient and hence variable” (e.g., CRISTÓFARO-SILVA; ALMEIDA, 2008).

As can be seen, some authors defend the variable nature of epenthetic vowels in BP (cf. CANTONI, 2015; MONARETTO, 2015; COLLISCHONN, 2004), whereas other scholars argue for the variability of epenthesis as a result of phonetic implementation (cf. JOHN; CARDOSO, 2017). Regarding the dialect of interest, a sociolinguistic study focusing on the acquisition of triconsonantal sequences with coronal plosives between heterosyllabic consonants by L2 English learners from the city of Campina Grande in the northeastern state of Paraíba, Brazil, shows interesting results (REIS; LUCENA, 2019). Overall, the analysis indicates that, out of a total of 1,071 tokens with the triconsonantal

sequence Ct/d]oC produced by 24 informants, 529 of them demonstrate the maintenance of the three-member consonant sequences, that is, 49% of the data were produced *without the use of any repair strategy* to break up this structure in L2 English speech, which is marked and non-existent in their L1.

In fact, some previous studies have already determined an association between the occurrence of Vowel Insertion (VI) in data from both Portuguese and English as an L2, revealing the incidence of a linguistic transfer from BP to the interlanguage (SCHNEIDER; SCHWINDT, 2010). In addition, Lucena and Alves (2010) verified the extent to which coda plosives are manifested in two Brazilian dialects from the South and Northeast regions, showing that the tendency to apply the vowel insertion rule in the dialect from the state of Paraíba (PB) is somewhat higher (with a factor weight<sup>2</sup> of 0.55), compared with the southern speech, whose weight disfavors the use of such a repair strategy (0.44).

Considering these possibilities of dealing with unlicensed coda consonants in BP, the purpose of this work was to verify how speakers from the city of Campina Grande (PB) produce the six plosive consonants /p, t, k, b, d, g/ in both medial and final codas of native and loan words. More specifically, our study aimed to answer the following questions:

1. How do BP speakers from Campina Grande (PB) deal with coda plosives in their native speech?; and
2. Which linguistic and/or extralinguistic factors significantly influence vowel insertion after coda plosives in this northeastern dialect?

This article is divided into four other parts, in addition to the introduction. Thus, the following section will discuss the variable process of CCW, which we consider to be active in the dialect used by BP speakers from the state of Paraíba, as already

---

<sup>2</sup>Factor Weight (FW) is a probabilistic measure whose values range from 0 to 1 (GUY; ZILLES, 2007). In the case of a binary analysis, the neutral point is equivalent to 0.5, which means that a factor has no effect on the use of the application value. Factor weights above 0.5 are interpreted as being favorable to the use of the application value, whereas any weight below the neutral point disfavors it.

demonstrated by Lucena and Alves (2010). Then, details will be provided regarding the methodological procedures adopted to achieve the formulated objectives. In the fourth section, we will explain the factors statistically significant for the insertion of a high front vowel after coda stops. Finally, we will point out some final considerations with possible contributions to the area of sociolinguistics.

### Surface forms of illicit consonants in BP codas

The post-vocalic position limits the possibilities of consonants, reducing the BP consonantal system to fricatives, nasals or liquids, represented by the archiphonemes /S/, /N/, /R/ and /L/, respectively. However, several words in BP reflect an old grammar (BISOL, 1999) and thereby contain coda obstruents, since voiceless stops were allowed in both Classical Latin and Vulgar Latin (VL) codas (CSER, 2016), as in VL “*raptāre*”, “*pactum*” and “*signum*” > BP “*raptar*”, “*pacto*” and “*signo*” (“to kidnap”, “pact” and “sign”, respectively). In fact, some of the consonant sequences<sup>3</sup> contained in the target words used in this study – as [pt] (“*capturar*”, “*adaptar*”), [kt] (“*detectar*”, “*aspecto*”, “*pacto*”), [ps] (“*corrupção*”, “*acepção*”), [ks] (“*flexão*”) and [gm] (“*dogma*”, “*enigma*”), for instance – are found in Classical Latin, based on the list formulated by Cser (2016). According to Collischonn (2004), however, some words containing coda obstruents entered the lexicon as a result of borrowing, since a few historical processes have led to the reduction or elimination of closed syllables in archaic Portuguese.

Although epenthesis is a distinctive strategy generally used to break up the contact between these heterosyllabic consonants, which leads to the pronunciation of the aforementioned Latin-derived words as “*rap[i]tar*”, “*pac[i]to*”, and “*sig[i]no*”, the incorporation of the unlicensed consonants into the coda is also a possibility in BP. Indeed, the epenthetic element seems to be specially avoided by educated speakers in more formal contexts (CAMARA JR., 1970/2013).

---

<sup>3</sup>We adopt the terminology proposed by Pulgram (1965), who distinguished the terms “cluster” and “sequence” with respect to the possible combinations of consonant sounds within words. According to the author, consonant clusters involve the co-occurrence of consonants within the same syllable. Consonant sequences, on the other hand, refer to the contact between consonants across syllable boundaries, i.e., word-internal coda-onset sequences.

As we have already seen, this variable phenomenon was investigated by Lucena and Alves (2010), who carried out a sociolinguistic research to verify, among other objectives, the incidence of CCW in the dialects from the southern state of Rio Grande do Sul and the northeastern state of Paraíba, as spoken by 22 undergraduate students enrolled in Language and Literature courses. Their informants were asked to read carrier sentences with words in both English and Portuguese containing /p/, /k/, and /f/ in codas followed by the coronal plosive /t/ in the subsequent syllable, as in “*optar*” (“to opt”), “*pacto*” (“pact”), “*afta*” (“canker sore”), “captain”, “doctor”, “after”, etc.

Their analysis shows that “[...] the dialect of Paraíba has a factor weight of 0.55 for the application of the epenthesis rule, whereas the southern dialect presents the value of 0.44” (LUCENA; ALVES, 2010, p. 38). Lucena and Alves (2010) acknowledge that there was no significant difference between the results of the two dialects, since an epenthetic element was added in 55.7% of a total of 253 tokens by the 12 southern speakers, compared with the relatively higher proportion of epenthesis in the northeastern speech (65.7% of 271 tokens). As can be seen, these rates of epenthesis indicate variability in the *corpus*. In the case of the southern dialect, vowel insertion was observed in 55.7% of the tokens, while the remaining 44.3% of the data demonstrate the incorporation of /p/, /k/, and /f/ into the coda by CCW. Although the rate of epenthesis in the northeastern dialect is slightly higher, it certainly should not be viewed as a categorical rule, considering that 34.3% of the tokens present the absence of an epenthetic vowel.

Therefore, we believe that this variable process is possibly active in the dialect from the state of Paraíba, especially in medial codas, as in the case of the following words: “*fac.ção*” [fak'sẽõ] ~ [faki'sẽõ] (“faction”); “*op.ção*” [ɔp'sẽõ] ~ [ɔpi'sẽõ] (“option”), “*pac.to*” ['paktu] ~ ['pakitu] (“pact”); and “*rép.til*” ['hɛptɪɫ] ~ ['hɛpitɪɫ] (“reptile”). Word-final plosives, however, tend to be mostly heard with a vowel insertion<sup>4</sup>, as occurs in several loanwords, such as “rock[ɪ]”, “pop[ɪ]”, “back-up[ɪ]” and “top[ɪ]”. In the following section, we explicit all the words used in the data collection instrument, as well as the steps adopted in conducting the research.

---

<sup>4</sup>The term “vowel insertion” is used with respect to the addition of phonetic substance in any word position, whereas the word “epenthesis” refers specifically to segments inserted word-medially.

## Method

This section presents a detailed description of the steps to perform the experiment, providing information on the research design, the statistical package used in data analysis, the informants, and the instruments applied in data collection.

### *Research design*

The present study is framed within the quantitative paradigm of sociolinguistic research pioneered by Labov (1963, 1966). Widely known as Variationist Sociolinguistics, this area may also be referred to as: (1) Labovian Sociolinguistics, in reference to the American linguist William Labov, whose research substantiated this new way of seeing the linguistic structure; (2) Quantitative Sociolinguistics, in view of “[...] operating with numbers and statistical treatment of collected data” (TARALLO, 2007, p. 8); and (3) Theory of Variation and Linguistic Change, focusing on the two main objects of study in this area. Each of these designations emphasizes a particular aspect of the theory, whose major empirical task is to “[...] correlate linguistic variation as the dependent variable with independent variables (CHAMBERS, 2003, p. 17 apud TAGLIAMONTE, 2012, p. 7). In the case of the present study, as we shall see in more detail in the Method section, we aim to systematize and order the constraints which lead to either the presence or absence of VI in the speech community of Campina Grande (PB).

### *Participants*

Data were collected with BP speakers born in Campina Grande, the second largest city in the state of Paraíba. A total of 12 informants were invited to participate in this study and, upon acceptance of the invitation, they were asked to fill in the Informed Consent Form, which provides information on the voluntary nature of their participation in the research as well as the line of research and graduate program to

which this work is linked. In addition, all the informants also answered the questionnaire, applied in both print and online versions<sup>5</sup>.

The conduction of this research with the subjects was duly approved by the by the Ethics in Research Committee, Health Sciences Center, Federal University of Paraíba (CEP/CCS-UFPB, Certificate of Presentation for Ethical Appreciation No. 70943917.3.0000.5188), thereby respecting and guaranteeing all ethical and deontological diligences considered essential for research involving human beings.

One of the participants was an undergraduate student at the time of data collection, while the rest of them had finished different undergraduate courses such as law, architecture, social work, geography, etc. They *had all studied English* in junior and senior schools for at least six years. Seven of them reported having also studied English in private language schools as well. Table 1 presents the informants' classification in the present study:

Table 1 – Informants' classification according to gender and L2 proficiency

| Coded names | Gender    | L2 proficiency levels | Ages |
|-------------|-----------|-----------------------|------|
| F1          | Cis women | Basic                 | 21   |
| F2          |           |                       | 55   |
| F3          |           | Intermediate          | 24   |
| F4          |           |                       | 24   |
| F5          |           |                       | 34   |
| F6          |           | Advanced              | 18   |
| F7          |           |                       | 31   |
| M1          | Cis men   | Basic                 | 33   |
| M2          |           | Intermediate          | 25   |
| M3          |           |                       | 42   |
| M4          |           | Advanced              | 32   |
| M5          |           |                       | 39   |

Source: Prepared by the authors

The statistical analysis of the 960 tokens will seek to shed light on the factors that condition the emergence or absence of VI in the speech of the 12 informants. In other words, we assume that the non-application of any repair strategies in syllables closed by

<sup>5</sup>Available online, the questionnaire applied in its digital version can be accessed through the site: <<https://docs.google.com/forms/d/e/1FAIpQLSekqcVAN5hbxOk3ZCv4W2L3awodbQITPKZIE5XinFhD-vMn6w/viewform>>.



plosives implies that the variable process of CCW is active in the speech of the population investigated, which is composed of individuals belonging to the socioeconomic strata from C1 to A, according to the Brazilian Economic Classification Criteria proposed by the Brazilian Market Research Association (2016).

### *Materials*

The group of 12 informants were asked to read aloud a list of 75 words embedded in the following carrier sentence: “The word is...”. To prevent informants from becoming aware of the structure in advance, the 75 crucial words were merged with 117 distracting words. Each of the 48 slides – exhibited on the screen of a laptop computer through a Microsoft PowerPoint presentation file (.ppt) – contains 4 carrier sentences, ending with words belonging to the same semantic field, such as food, sports, computer science, politics, professions, feelings, family, etc.

Regarding the organization and display of these words more specifically, each slide exposed in the PowerPoint application shows four carrier sentences, with at least one of them containing a crucial word, while the remaining phrases are closed with distracting words, all within the same semantic field as the target word, as previously mentioned. For example, the three carrier phrases displayed on the same slide as the word “*pacto*” (“pact”) are: (1) “The word is *Acordo*”; (2) “The word is *Convenção*”; and (3) “The word is *Tratado*”. By applying the animations feature, available on the standard PowerPoint application toolbar, the carrier phrases could be displayed one at a time, only requiring a left click to appear sequentially on the screen, thereby ensuring that the sentences were read according to the rhythm of each participant.

The 75 crucial words containing syllables closed by plosives include both lexical items which are etymologically categorized as “erudite” – because they entered the Portuguese lexicon from written language as late loans from classical Latin (CAMARA JR., 1970/2013; BIZZOCCHI, 2013) – and loan words. These 75 target words embedded in 192 carrier sentences, all of which contain coda plosives in word-medial (as in 44 items)

and -final positions (in 31 words), are divided into 45 native words (in italics) and 30 loan words commonly used in BP. All the lexical items are listed in Table 2:

Table 2 – Target words used in the data collection instrument

| Stray consonant (C') | Word-medial position  | Word-final position                                       |
|----------------------|---|---|
| /p/                  | <i>corrupção; opção; adaptar; capturar; acepção; laptop; Snapchat; Flipboard</i>    | <i>USP; WhatsApp; top</i>                                 |
| /b/                  | <i>obter; óbvio; subversivo; abdominal; abstrair; abstrato; obstrução; observar</i> | <i>sob; kebab; web; tab; Job; UNILAB</i>                  |
| /t/                  | <i>etnia; Vietnã; atmosfera; ritmo; futsal; quartz; Netflix; Jetsons</i>            | <i>habitat; tablet; chat; internet; post; site</i>        |
| /d/                  | <i>advogado; administrar; advertência; admitir; adjunto</i>                         | <i>Unimed; Harvard; Oxford; Flipboard; iPad; download</i> |
| /k/                  | <i>detectar; aspecto; pacto; tecnologia; cacto; bactéria; fictício; flexão</i>      | <i>PUC; Facebook; Big Mac; like; esc; park</i>            |
| /g/                  | <i>ignorante; significado; signo; dignidade; Bagdá; dogma; enigma</i>               | <i>VARIG; Burger King; hot dog; blog</i>                  |

Source: Prepared by the authors

Some of the loan words listed in Table 1, such as “laptop”, “Flipboard”, “hot dog”, “Snapchat”, and “Big Mac”, contain two syllables closed by plosives, which means that 5 extra syllables were captured in addition to the 75 of each crucial word, totaling thereby 80 tokens per participant. Consequently, the *corpus* is composed of 960 outputs, which were recorded in audio through the software Audacity, v. 2.0.6 (MAZZONI, DANNENBERG et al., 2014).

## Data Analysis

In this section we start by discussing the statistical software and methods used in data analysis and then concentrate on the *factors with statistically significant effects on the dependent variable (response) of interest, whose application value is VI.*

### *The Rbrul software package*

The reflexes of technological advances in the evolution of statistical packages for the quantitative treatment of variable language data are undisputed. Since the beginning of variationist studies, researchers have employed different mathematical models (e.g., additive model, multiplicative model and logistic models) underlying the computer programs for multivariate analysis over the last 45 years (GOMES, 2012).

In general, the different tools developed for this function seek to identify which factors influence the use of a binary linguistic variable (necessarily conceived in terms of opposition between two discrete alternatives, such as the presence or absence of a given linguistic element), in addition to measuring the effects of such factors on the conditioning of variants (JOHNSON, 2009). Factors conditioning the variation may be associated with the language structure itself – internal factors, such as phonological contexts, stress, morphological or syntactic classes, etc. –, and with the use that different individuals make of it in a given speech community – external factors, such as age, gender, ethnicity, socio-economic background, level of education, etc. (GOMES, 2012). Regarding these variables more specifically, Johnson (2009) points out that:

Fixed effects are factors with a fairly small number of possible levels, for example, male/female, stressed/unstressed, or following vowel/consonant/pause. These factors are usually the direct object of interest, and their levels would be replicable in a further study (JOHNSON, 2009, p. 364-5).

There are factors, however, that are drawn from truly larger populations which are not easily replicable in other studies, such as individual speakers or specific words (JOHNSON, 2009). Thus, even when speakers or words are incorporated into fixed-effect models, these variables are often treated as fixed, which may lead to the attribution of the effect to the factor group as a whole. Unlike previous variable rule programs like VARBRUL (CEDERGREN; SANKOFF, 1974) and GoldVarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005), the Rbrul statistical package can compute both fixed-effects and mixed-effects models, since it can perform a binary logistic regression with both fixed and random effects, i.e., factors that are nested within other factors.

Another crucial difference between Rbrul and its predecessors involves the use of discrete categories in software for fixed-effects models, and both discrete (categorical)

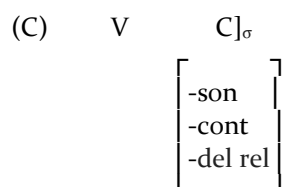
and continuous variables in the Rbrul package. This means that variables like age or lexical frequency, for example, can be handled as continuous predictors.

As regards the statistical analysis performed by the software, Hosmer and Lemeshow (2000, p. 1) claim that “[r]egression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable [dependent variable] and one or more explanatory variables [independent variables].” An overview of the variables controlled in this study will be provided in the following section.

### Variables

The phonotactic constraints of BP are likely to explain the fact that when a plosive occurs in coda position, it is common to use an epenthetic vowel on the surface, in words such as “ad[i]vogado”, and “ad[i]mirar”, thereby breaking up illegal consonant sequences across syllable boundaries by inserting an extra vowel, so that /d/ is *resyllabified into the onset of the newly-created syllable*. This is precisely what we aim to investigate in this paper: the extent to which VI occurs after coda plosives in the dialect of Paraíba. Figure 1 portrays the syllable structure of interest:

Figure 1 – Syllable structure of words from the instrument



Source: Prepared by the authors

Table 3 presents the predictors of VI after coda plosives included in the logistic regression model:

Table 3 – Summary of factor groups considered in Rbrul

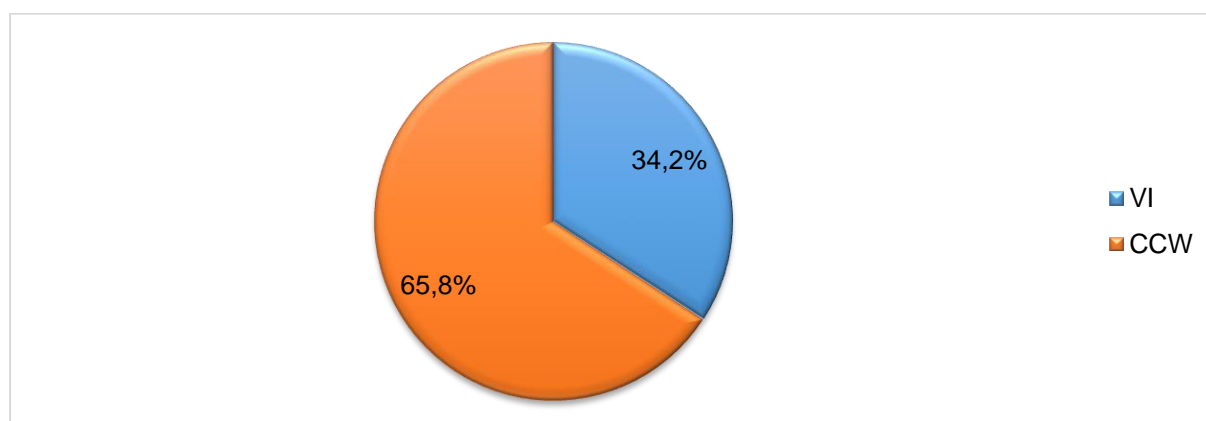
| Predictors                 | Levels                                   | Type        |
|----------------------------|--|-------------|
| Informants                 | M1-M12, H1-H12                           | Random      |
| Words                      | <i>Corrupção, opção, adaptar</i> etc.    | Random      |
| Age                        | Min. 18; median 32; mean 34,183; max. 60 | Continuous  |
| Gender                     | Cis men and cis women                    | Categorical |
| L2 proficiency levels      | Basic, intermediate, and advanced        | Categorical |
| Stray consonants           | /p/, /t/, /k/, /b/, /d/, /g/             | Categorical |
| <i>Etymological status</i> | Native and loan words                    | Categorical |
| Word size                  | Min. 1; median 2; mean 2.388; max. 5     | Continuous  |

Source: Prepared by the authors

### Rbrul results for VI after coda plosives

Figure 2 presents the rates of VI and CCW in the 960 productions of coda plosives by the 12 BP speakers from the city of Campina Grande:

Figure 2 – Overall frequency of vowel insertion after coda plosives



Source: Prepared by the authors

As can be seen, plosive phonemes in coda position surfaced variably in the dialect of Paraíba: 328 tokens present an additional vowel, whereas the plosives were syllabified in their original position in the other 632 tokens. This means that a specific group of informants resorted to VI in some words more than others (cf. Table 1), including

“Unimed” (the name of a [Brazilian health insurance](#) company), “Bagdá” (“Bagdad”), “dignidade” (“dignity”), “etnia” (“ethnicity”), “sob” (“under”), “administrador” (“administrator”), “admitir” (“to admit”), “adjunto” (“adjunct”), “advertência” (“warning”), “Varig” (the name of the first airline founded in Brazil), and “enigma” (“enigma”), which were pronounced with additional phonetic substance in *all* their respective 12 tokens. However, words like “futsal” (“five-a-side football”), “quartzô” (“quartz”), “acepção” (“meaning”), “flexão” (“push-up”), “tablet”, “snapchat”, “post”, “laptop”, “chat”, “park”, “like”, “oxford”, and “download”, for example, were widely produced without vowel insertion. Naturally, these are only preliminary observations: in order to better understand the phenomenon of VI after coda plosives, we need to interpret the results of the statistical analysis.

The factors selected as statistically significant for the occurrence of VI in the dialect from Paraíba are: gender + etymological status + L2 proficiency + stray consonant. They have a *p*-value of less than 0.001, and the results obtained for each factor group will be discussed separately in the following sections.

### Gender

The effects of “gender” are quite relevant to explain the phenomenon in the speech community under analysis. VI was detected in only 22% of the 400 tokens with illicit coda consonants produced by male speakers, compared to women's rate, which is almost double that of men: 43% of the 560 outputs present additional phonetic substance. Table 4 provides the logistic regression results of gender:

Table 4 – Effect of gender on the emergence of VI

| Factors | Log-odds | Cases/Total | %  | Factor Weight |
|---------|----------|-------------|----|---------------|
| Women   | 0.649    | 240/560     | 43 | .66           |
| Men     | -0.649   | 88/400      | 22 | .34           |

Deviance: 913.718; *p* < 0.001; df: 9; Intercept: -1.228; Input probability: 0.227; Grand mean: 0.342

Source: Adapted from the Rbrul's output

With an effect size<sup>6</sup> of 32 – which is considered a strong estimate according to the taxonomy outlined by Horvath and Horvath (2003 apud ROY; OETTING; MOLAND, 2013) to explain the range –, the seven women inserted vowel segments into the words almost twice more (43%) than the men (22%). These results may reflect differences in both *behaviors* and *attitudes* towards the English language. The participants H3 (only in some cases), H4 and H5 (more conspicuously) pronounced the foreign words with their original English sounds, using the retroflex approximant [ɹ] and *the aspirated allophones of voiceless stops* ([p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>]) in stressed syllables, for example. Curiously, none of the female participants did so. Only two male subjects produced the directly-imported English words with the *adaptation of the foreign sounds to the native phonological system* (“internet” pronounced as [ĩtɛh'netɪ], for example).

The data in Table 5 provide further details about the different behaviors, derived from a cross-tabulation of the gender factor in relation to the etymological status of the words:

Table 5 – Cross-tabulation of gender and etymological status effects on the surface forms of coda plosives

| Gender →<br>Etymology ↓ | Levels | Cis women |    | Cis men |    |
|-------------------------|--------|-----------|----|---------|----|
|                         |        | N         | %  | N       | %  |
| Native words            | VI     | 191       | 59 | 70      | 30 |
|                         | CCW    | 131       | 41 | 165     | 70 |
|                         | Total  | 322       |    | 235     |    |
| Loan words              | VI     | 49        | 21 | 18      | 11 |
|                         | CCW    | 189       | 79 | 147     | 89 |
|                         | Total  | 238       |    | 165     |    |

Source: Prepared by the authors

Table 5 shows that 21% of the 238 loan words produced by the female participants contain an epenthetic vowel. Thus, the percentage of women using the VI in foreign words is 10 points higher than that of the male informants, who added a vowel segment

<sup>6</sup>The range or effect size is reached by subtracting the lowest factor weight from the largest. The larger the absolute value, the stronger the effect. More specifically, Horvath and Horvath (2003 apud ROY; OETTING; MOLAND, 2013) distinguish the following classifications: (1) an effect size of less than 10 is considered a weak one; (2) values between 10 and 30 suggest a moderate effect; (3) between 30 and 50, the effect is considered strong; and finally, (4) a value greater than 50 represents a very strong effect.

in only 18 out of the 165 imported English words (11%). Regarding the use of VI in native words, out of the total of 322 tokens, the group of women inserted a front vowel segment in 191 tokens (59%), whereas the men added phonetic substance in only 30% of the 235 words belonging to the PB lexicon. As can be seen, the data show a greater behavioral difference in the use of VI after coda plosives in the context of lexical items from BP, although the rates of VI are higher in women's speech than that of men, considering both native and loan words produced by the participants. As a result, women seem less comfortable with coda plosives for reasons related to social identification.

Given the results discussed here, we infer that the transformation of the closed syllable structure into two open syllables (CVC → CVCV) also reflects identity processes that go beyond a deterministic and fixed classification of subjects in terms of their membership to certain social, political and/or cultural groups, including that of gender (ENNES; MARCON, 2014). Based on Ennes and Marcon's (2014, p. 274) proposal, we adopt the notion of identity "[...] as a dynamic and in-process social phenomenon, fundamentally implied by power relations", and thereby we understand genders as markers that are continually constructed through social relations in particular socio-historical contexts.

Considering this view of identity processes formulated by Ennes and Marcon (2014), we assume that the use of phonetic substance, which refers to a typical repair strategy of BP, reflects dynamic demarcations in the construction of the social boundaries between I/we and the others. The contrast between the presence and absence of vowel insertion is, in our view, socially marked, considering how social actors construct the boundaries between what they understand as "we" and "them", based on the sense of belonging, interests, wishes and their individual and collective experiences as well (ENNES; MARCON, 2014). Therefore, the use of VI by the female participants reinforces their local identity, by simplifying syllables closed by plosives, which are common in English, but illegal in BP.

Women tended to apply the phonological rule  $\emptyset \rightarrow i / [-\text{son}, -\text{cont}] \_\_\_\_\_\_ ]_\sigma$  more often than men, especially in the lexical item "*Vietnã*" ("Vietnam"), which was produced by all the female informants with the insertion of a high front vowel, whereas all the male participants produced the same word maintaining [t] in coda through the CCW.



Other items, such as “*adaptar*” (“to adapt”) and “*fictício*” (“fictitious”), were also pronounced by most women with the addition of phonetic substance: in the case of these two words specifically, there were only two occurrences of each produced by the seven female speakers without epenthesis.

### *Etymological status*

The etymological status of the words had a great impact on the process of *vowel insertion in words* containing coda plosives. With a strong effect size of 42, the English words did not trigger the application of VI in the same way as native words, with factor weights of 0.29 vs. 0.71, respectively. Table 6 shows the results concerning the effects of the etymological status on VI:

Table 6 – Effect of words’ etymology on VI

| Factors | Log-odds | Cases/Total | %  | Factor Weight |
|---------|----------|-------------|----|---------------|
| Native  | 0.896    | 261/557     | 47 | .71           |
| Loan    | -0.896   | 67/403      | 17 | .29           |

Deviance: 913.718;  $p < 0.001$ ; df: 9; Intercept: -1.228; Input probability: 0.227; Grand mean: 0.342

Source: Adapted from the Rbrul’s output

These results show us that there is some resistance in the addition of phonetic substance to the English words. The fact that all the participants use English as an L2 may explain why the loan words tended to be pronounced with foreign sounds. This observation is in agreement with the suggestion proposed by Pike and Fries (1972, p. 62) that one of the differences between monolinguals and bilinguals is that speakers of only one language are more susceptible than those of two or more languages “[...] to force the loan words to conform to the native phonetic and phonemic pattern.” In contrast, nearly half of the total 557 native tokens present the creation of an additional syllable to save the stray consonants.

As a comparison, the acronym “*Unimed*”, for example, was produced by all the informants with the addition of [ɪ] after /d/ in word-final position, whereas the word

“download” (also ending in /d/) was not resyllabified in none of its 12 occurrences. It is possible that the vowel in the word “*Unimed*[ɪ]” is phonological and is part of the lexical representation. Indeed, Collischonn (1996) considers epenthesis to be a lexical and not only post-lexical process. One of the arguments put forward by the author can be applied to this acronym: epenthesis is viewed as a lexical phenomenon because it is maintained across word boundaries as in, for example, “*a Unimed*[ɪ] *oferece* planos individuais” (not \**a Unimed*[ɔ] *ferece* planos individuais or “Unimed offers individual plans”). In this case, resyllabification should block vowel insertion, but that does not occur.

Moreover, this demonstrates that individuals seem to distinguish the coda conditions of the two languages or, perhaps, the most prominent phonotactic constraints, such as those governing the segments which may occur word-finally in Portuguese, namely the vowels – both stressed and unstressed – or the archiphonemes /L, R, N, S/, and in English, whose words can end in reduced or tense (long) vowels (AKAMAJIAN et al., 2010), as well as in any consonant from the inventory, except [h].

### *L2 proficiency*

With respect to L2 proficiency, all subjects *had studied English* for *at least* 6 years in secondary schools, and some of them had also studied in private language schools. Their proficiency was divided into three levels, namely: basic, intermediate and advanced. This classification is widely adopted in sociolinguistic studies in general and was determined based on proficiency test scores. Thus, the 12 informants were asked to answer 200 multiple-choice questions from the Oxford Placement Test (ALLAN, 2004), which is an English proficiency test validated in more than 30 countries. Table 7 provides details on how the 12 informants were classified into the three levels using their Oxford test scores:

Table 7 – Proficiency levels calculated from scores obtained on test

| Scores | L2 English proficiency levels |
|--------|-------------------------------|
| 0-119  | Basic                         |

|         |              |
|---------|--------------|
| 120-149 | Intermediate |
| 150-200 | Advanced     |

Source: Adapted from Dabbagh and Noshadi (2014, p. 236)

In Portuguese, there are no native words ending with the voiced coronal plosive, even if it is possible to find some words ending in /d/, including: (1) some proper names (“Astrid”, “David”, “Ford”, “Fred”, “Haddad”, “Mohammed”, “Said”, etc.); (2) some native and foreign acronyms, and abbreviations (such as “Unimed”, “Unicred”, “Hipercard”, “LED” and “TED”); and clearly, (3) foreign words (“iPad”, “iPod”, “download”, etc.). The native dialect used by the BP speakers seems to be affected by their interlanguage, in that the variable rule of VI, i.e.,  $\emptyset \rightarrow i$  / [-son, -cont] \_\_ ]<sub>σ</sub>, is slightly more applied by elementary L2 learners. Table 8 shows Rbrul’s output:

Table 8 – Effect of proficiency levels on VI

| Factors      | Log-odds | Cases/Total | %    | Factor Weight |
|--------------|----------|-------------|------|---------------|
| Intermediate | 0.204    | 150/400     | 37.5 | .55           |
| Basic        | 0.093    | 89/240      | 37   | .52           |
| Advanced     | -0.297   | 89/320      | 28   | .43           |

Deviance: 913.42;  $p < 0.05$ ; df: 10; Intercept: -1.161; Input probability: 0.238; Grand mean: 0.342

Source: Adapted from the Rbrul’s output

Regarding the values of the “intermediate” and “basic” factors, it should be noted that there is a very subtle difference between the percentages and the factor weights of the two groups: while intermediate learners used VI in 37.5% of the 400 tokens with coda plosives, speakers at elementary L2 levels present a slightly lower application rate of 37%. The values obtained for these two groups demonstrate that both intermediate and basic learners – with factor weights of 0.55 and 0.52, respectively – favor the application of VI in a similar manner. In view of these comparable proportions, we decided to verify if the combination of basic and intermediate speakers into a single category leads to a more accurate model.

According to Daleszyska (n.d.), to compare different runs, we need to apply the log likelihood-ratio test, whereby the log-likelihood values from both runs are calculated to help determine which model offers a better fit to the data. By dividing the deviance

values from both models by -2, we achieved the following log likelihood values for the runs with three L2 proficiency factors (basic, intermediate, and advanced) and two factors (elementary and advanced), respectively: (1) -456.71; and (2) -456.859. We now subtract the log likelihood of the large model from that of the small one, and multiply the result by 2 in order to obtain the deviance between them:  $2(-456.71 - [-456.859]) = 0.298$ . Given that we have three factors in the first run and two in the second one, the difference in degrees of freedom (df) between them is 1, and in the chi-square table, at 1 df (left column) and  $p = 0.05$ , the critical value is 3.841, as can be verified in Table 9:

Tabela 9 – Chi-square distribution table

TABLE IV. DISTRIBUTION OF  $\chi^2$

Probability.

| <i>n</i> | .99   | .98   | .95   | .90   | .80   | .70   | .50   | .30    | .20    | .10    | .05    | .02    | .01    | .001   |
|----------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|
| 1        | .0157 | .0268 | .0393 | .0558 | .0842 | .148  | .455  | 1.074  | 1.642  | 2.706  | 3.841  | 5.412  | 6.635  | 10.827 |
| 2        | .0201 | .0404 | .0503 | .0711 | .1046 | .173  | 1.386 | 2.408  | 3.219  | 4.605  | 5.991  | 7.824  | 9.210  | 13.815 |
| 3        | .0115 | .0285 | .0352 | .0584 | .0905 | .1424 | 2.366 | 3.665  | 4.642  | 6.251  | 7.815  | 9.837  | 11.345 | 16.266 |
| 4        | .0297 | .0429 | .0511 | .0764 | .1069 | .195  | 3.357 | 4.878  | 5.989  | 7.779  | 9.488  | 11.668 | 13.277 | 18.467 |
| 5        | .0554 | .0752 | .0845 | .1100 | .1343 | .200  | 4.351 | 6.064  | 7.289  | 9.236  | 11.070 | 13.388 | 15.086 | 20.515 |
| 6        | .0872 | .1134 | .1235 | .1504 | .1770 | .248  | 5.348 | 7.231  | 8.558  | 10.645 | 12.592 | 15.033 | 16.812 | 22.457 |
| 7        | .1239 | .1564 | .1676 | .1933 | .2222 | .301  | 6.346 | 8.383  | 9.803  | 12.017 | 14.067 | 16.622 | 18.475 | 24.322 |
| 8        | .1646 | .2032 | .2133 | .2390 | .2704 | .352  | 7.344 | 9.524  | 11.030 | 13.362 | 15.507 | 18.168 | 20.090 | 26.125 |
| 9        | .2088 | .2532 | .2635 | .2898 | .3210 | .403  | 8.343 | 10.656 | 12.242 | 14.684 | 16.919 | 19.679 | 21.666 | 27.877 |
| 10       | .2558 | .3059 | .3160 | .3423 | .3735 | .457  | 9.342 | 11.781 | 13.442 | 15.987 | 18.307 | 21.161 | 23.209 | 29.588 |

Source: Adapted from Fisher and Yates (1963, p. 47)

Consequently, since the chi-square value (0.298) is lower than the critical value (3.841), it should be assumed that the difference between basic and intermediate speakers is not significant and, therefore, should not be maintained. In fact, Guy and Zilles (2007) argue for models with fewer factors by pointing out that:

Amalgamation analysis therefore uses fewer factors (which is always good according to Occam's razor), without losing significant information and without significantly worsening the quality of the approximation between the model and the data (measured by log likelihood) (GUY; ZILLES, 2007, p. 194).

Table 10 presents the effects of the two L2 proficiency factors of the simpler model on VI:

Table 10 – Effect of proficiency levels on VI

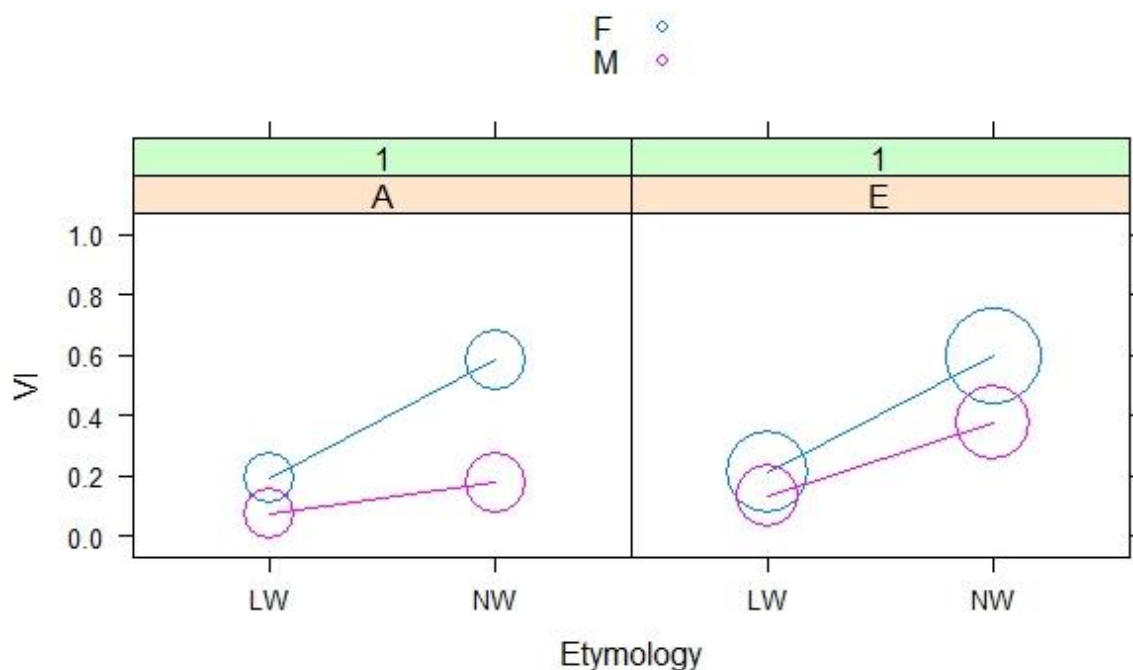
| Factors    | Log-odds | Cases/Total | %  | Factor Weight |
|------------|----------|-------------|----|---------------|
| Elementary | 0.23     | 239/640     | 37 | .56           |
| Advanced   | -0.23    | 89/320      | 28 | .43           |

Deviance: 913.718;  $p < 0.001$ ; df: 9; Intercept: -1.228; Input probability: 0.227; Grand mean: 0.342

Source: Adapted from the Rbrul's output

The results show that the advanced L2 speakers do not favor VI, with a FW of 0.43, considering that 72% of 320 tokens with [-son, -cont] ]<sub>σ</sub> surfaced without the application of any repair strategy to readjust them. Having already discussed three independent variables (IV), we now turn to the interrelationship between them. Figure 3 presents the plot of the relationship between three variables:

Figure 3 – Relationship between etymology, L2 proficiency, and gender



Source: Rbrul's output

This scatter plot shows how many tokens female (F) and male (M) speakers (IV “gender”, color) with different L2 proficiency levels (where A = advanced and E = elementary users, panels) produced using VI (dependent variable, y-axis) in native (NW)

and loan words (LW) (IV “etymology”, x-axis). The circles on the graph indicate the number of tokens, so that the bigger they are, the more data they represent. It is possible to observe, then, that the elementary female speakers inserted more vowels in native words than in loan words. Moreover, female speakers at advanced L2 proficiency levels used VI in native words at a higher rate than that of the men at the same L2 level.

In the following section we discuss the last statistically significant variable.

### *Stray consonants*

The Rbrul output shows that the voicing of the stray consonant promoted VI. Table 11 shows the results of the mixed-effects logistic regression analysis:

Table 11 – Effect of stray consonants on VI

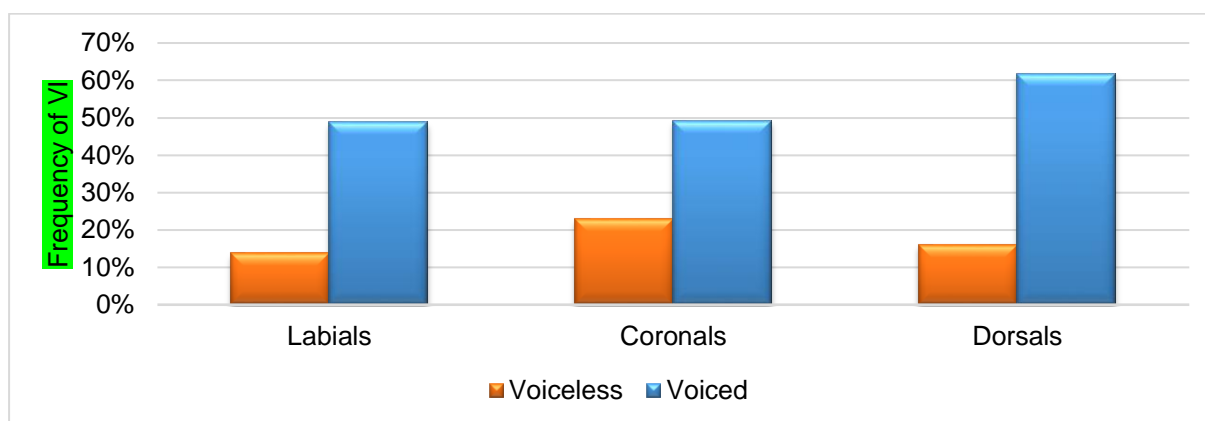
| Factors | Log-odds | Cases/Total | %    | Centered<br>Factor Weight |
|---------|----------|-------------|------|---------------------------|
| [g]     | 1.325    | 89/144      | 62   | .79                       |
| [d]     | 0.910    | 65/132      | 49.2 | .71                       |
| [b]     | 0.698    | 82/168      | 48.8 | .67                       |
| [t]     | -0.373   | 44/192      | 23   | .41                       |
| [k]     | -1.221   | 25/156      | 16   | .23                       |
| [p]     | -1.339   | 23/168      | 14   | .21                       |

Deviance: 913.718;  $p < 0.001$ ; df: 9; Intercept: -1.228; Input probability: 0.227; Grand mean: 0.342

Source: Adapted from the Rbrul's output

As we can see, the plosives that favored VI are voiced, with factor weights of: 0.79 for the dorsal; 0.71 for the coronal; and 0.67 for the labial. Their voiceless counterparts, on the contrary, are shown to inhibit vowel insertion, so that these sounds tended to be realized on the surface in the original coda position, as shown by the low values of factor weights: 0.41 for [t]; 0.23 for [k]; and 0.22 for [p]. Figure 4 displays the effects of the [+voice] segments on VI, contrasting the rates between voiced/voiceless pairs of plosives:

Figure 4 – Distribution of VI by place of articulation and voicing contrast



Source: Prepared by the authors

It is clear from Figure 4 that the proportion of VI after voiceless plosives is quite low, not reaching  $\frac{1}{4}$  of data in any of three segments. The frequency of the application value after voiced plosives, on the other hand, is markedly higher. In the case of [b] and [d], front vowels were inserted in more than double the amount of their respective counterparts. As for [g], the incidence rate is almost four times higher than that of [k]. These results corroborate the assumption that the emission of a vowel segment would be facilitated by the vocal fold vibration with an effect size considered very strong, according to the bands of robustness proposed by Horvath and Horvath (2003 apud ROY; OETTING; MOLAND, 2013): out of the six consonants investigated in the coda position, VI is promoted more intensely by the voiced stops /g, d, b/ respectively, in words like “si.g[ɪ].no”, “VA.RI.G[ɪ]”, “a.d[i].ver.tên.ci.a”, “U.ni.me.d[ɪ]”, “o.b[i].ter”, and “so.b[ɪ]”, for example.

Taking into account only the percentages related to the voiceless stops in Table 11, we see a rather interesting fact: there is a slight difference between the application of VI when C' is [t] or [k] and [p]. Although the factor weights (FW) of less than 0.5 indicate that these voiceless stops disfavor the insertion of vowels, the voiceless coronal stop has a percentage and a FW slightly higher than those of the non-coronal segments. Thus, words such as “ritmo”, “atmosfera”, and “internet” were resyllabified a little more often than items containing /k/ or /p/ in coda, such as “aspecto”, “pacto”, “PUC” and “adaptar”. The word “opção”, for example, was produced only once as “o.p[i].ção” (by the informant

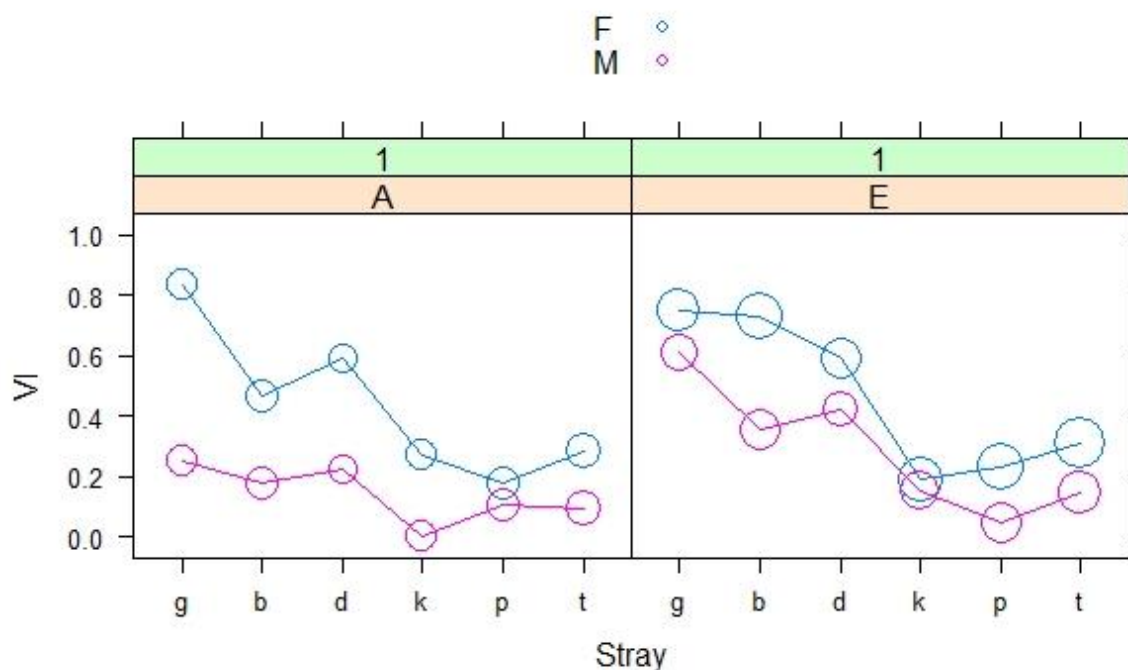
F3); in the other 11 phonetic realizations of this item the plosives were maintained in coda position by CCW. Another example illustrates how /p/ and /k/ were less resyllabified than /t/: the voiceless coronal plosive in the verb “*adaptar*”, with the marked consonant sequence /p.t/, was resyllabified in only two of the 12 tokens, as produced by F4 and F7. Figure 4 shows that violations of the coda condition in our sample were favored by the voiceless stops /p/, /k/, and /t/, respectively.

It follows that codas closed by plosives tend to have the Brazilian Portuguese coda condition weakened in the dialect of Paraíba, allowing the underlying obstruents to surface in the original position of coda, as shown in 632 out of the 960 tokens collected. In the other 34.2% of the occurrences, VI was used, favored by vocal pulsation at the margin of the closure interval. In fact, the velar stops have longer voice onset time (henceforth VOT) values than labials, which have the shortest VOTs, and coronals, with intermediate VOT values (KENT; READ, 2015). Considering the universal nature of VOT interaction with the consonant’s point of articulation – with smaller values for more anterior constriction locations (KENT; READ, 2015) –, it is possible to observe that the order established here may be related to this phenomenon.

In order to explore all these variables in more detail, *the plot in* Figure 5 shows how female (F) and male (M) speakers (IV “gender”, colors) at different L2 proficiency levels (where A = advanced and E = elementary speakers, panels) resorted to VI depending on the stray consonant (SC, x-axis):



Figure 5 – Relationship between L2 proficiency, gender and stray consonants



Source: Adapted from the Rbrul's output

Again, this scatter plot clearly shows that women (blue line) in the elementary and advanced proficiency groups inserted vowels more frequently than the male informants, especially after the voiced dorsal plosives, in words like “*dignidade*” (“dignity”), “*enigma*” (“enigma”), “*admitir*” (“to admit”), “*advogado*” (“lawyer”), “*abdominal*” (“abdominal”), and “*óbvio*” (“obvious”), for example. In view of the results discussed here, it is manifest that there is a clash between the two options in the population investigated: although women have higher rates of the application value, compared with those of men, 57% of all the 560 tokens produced by the female group have surface forms violating Brazilian Portuguese coda condition.

In our view, women seem to resort more often to resyllabification not only to readjust an ill-formed syllable structure, but rather to convey identity information and social status. As previously discussed, some of the male participants produced the anglicisms with an Americanized pronunciation, which may indicate an inclination or openness of these individuals to US values or to the presence of American culture in the

country, in the form of products disseminated by the cultural industry through music, cinema, fashion and TV (MADEIRA, 2002).

The presence of VI is strongly favored by the group of women in our investigation, and it is crucial to understand why this is the case by addressing the social meaning in language use. In order to do so, we need to reflect on the current global status of the English language.

According to Crystal (2003, p. 9), “[a] language has traditionally become an international language for one chief reason: the power of its people – especially its political and military power.” He points out that this kind of power plays a decisive role in establishing a language as an international medium of communication, but it takes more than military might to maintain and expand that language. These latter tasks can only be pursued through cultural and economic factors. Therefore, the maintenance of English in its current position as a global lingua franca is the result of the economic supremacy of the United States, as well as its cultural power throughout the world, in addition to its political leadership and technological domination.

We agree with Silva’s claim that “[t]he cultural tastes and preferences of Brazilian society strongly reflect American cultural domination” (2015, p. 22), in the face of the plethora of empirical and incidental evidence. Although women may share the same interests that we attribute to men in our analysis, none of them relied on the English sounds to pronounce the foreign words introduced into our language. In addition, we also assume that the more frequent use of coronal vowels to repair the more marked (also considered ill-formed) codas can be taken as demarcation of their local identity, in contrast to the global identity associated with the English language nowadays (GREY; PILLER, 2019).

In short, the status of English around the world as a global lingua franca, as well as the role played by globalization – as a process that drives US cultural and economic neo-imperialism (LOPES, 2006), whether through the exchange of information and the sharing of markets, or the inevitable cultural “exchange”, which is continually facilitated and disseminated by the rapid advance in technology – are crucial aspects to understand the development of linguistic forms of resistance.

The selection of gender demonstrates that variable patterns cannot be fully explained only through the perspective of language structure per se, without taking into consideration how language is intertwined with culture. In this paper, the intimate relationship between individuals, language and society seemed to be reflected on the pronunciation of plosives in coda position, in that women born in Campina Grande prefer to avoid illegal coda consonants by inserting an additional vowel to construct a canonical syllable type. This reveals a possible form of resistance in the face of the relations of domination and power that permeate the current uses of English worldwide.

### Final remarks

The variable pronunciation of words like “*apto*”, “*ritmo*”, “*pacto*”, “*obter*”, “*adjunto*” and “*dogma*” was investigated in this work, which aimed to investigate the production of coda plosives in the dialect of Paraíba, by analyzing the factors which condition the presence or absence of Vowel Insertion after such sounds. A total of 960 occurrences of the phenomenon produced by 12 BP speakers from Campina Grande were recorded in audio and submitted for multivariate analysis using Rbrul (JOHNSON, 2015).

Two groups of linguistic factors and two groups of extralinguistic constraints were selected by the program as being relevant for the conditioning of VI after codas closed by plosives. The four independent variables analyzed here were: gender; etymology; L2 proficiency; and stray consonant. In view of all the results interpreted in this paper, we point out that codas closed by voiceless plosives tended to be phonetically realized without the application of *any repair strategy* in the speech of the group of BP speakers from Campina Grande (PB). Therefore, in 632 out of the 960 tokens (65,8%) the underlying plosives were maintained in the coda position on their surface phonetic representation. In the other 34.2% of the occurrences, VI was applied mostly by female speakers at both levels of L2 proficiency, especially in the production of native words with codas closed by voiced plosives.

The selection of gender allowed us to better understand the variability in the use of VI, revealing a relevant distinction between the behavior of men and women from the

city of Campina Grande. In general, we found that women at elementary and advanced L2 proficiency levels have higher rates of VI in both native and loan words containing codas closed by voiced plosives. Although the application value in directly-imported English words was low (in less than  $\frac{1}{4}$  of the data of both men and women), we observed that the rate of women adding phonetic substance to foreign words is 10 percentage points higher than that of men, who inserted front vowels in only 18 out of the 165 English words (11%): the female participants did so in 21% of the 238 loan words. Regarding the use of VI in native words, women once again demonstrated favoring such a repair strategy, since they applied it in 191 (59%) of the 322 words belonging to the BP lexicon, whereas the men added phonetic substance in only 30% of the 235 native words.

Overall, our results suggest that the female northeastern speakers seem to express their local and cultural identity by avoiding coda plosives on the surface form more strongly than the men, who demonstrate a positive attitude towards the American culture in general. Consequently, the use of VI has a social meaning: even though all the participants are users of L2 English, only some of the men demonstrated receptiveness to this global language when pronouncing the foreign words without *adaptation to the native* sound system. Considering the view of identity “[...] as socially constituted, a reflexive, dynamic product of the social, historical and political contexts of an individual’s lived experiences” (HALL, 2002, p. 31), we defend that the gender differences in this study may reflect contrasting attitude towards English, *but further investigations are needed to confirm this assumption*.

We are led to conclude that in addition to readjusting a marked structure – since closed syllables occur less frequently in BP, especially those closed by plosives –, the transformation of the CVC structures into two open syllables (CVCV) also serves as a social marker, which is continuously constructed in the process of social relations in certain socio-historical contexts, instead of being a fixed or static notion (ENNES; MARCON, 2014). Therefore, the use of a typical repair strategy in BP by the group of women investigated in our study may reflect a form of resistance regarding the American culture.

The dominant positions of English as an international language around the world and of men in our society may explain why the male informants were much more

comfortable with a structure that on the one hand violates the coda condition in BP, but on the other hand is allowed in English: they may perceive [-son, -cont] ]<sub>σ</sub> as having more prestige, possibly because it resembles a typical English pattern. In fact, some of the men produced the loan words with an Americanized pronunciation, which might suggest an interest in or an admiration for the *American culture*, society, and language.

Finally, by adopting an integrated approach to explain the linguistic variable of interest, the findings achieved here demonstrate that syllable boundaries readjusted with the creation of a new syllable structure in the words containing coda plosives also function as social boundaries, demarcating the Brazilian and individual identity of the female speakers from the city of Campina Grande.

## References

- ABEP – Associação Brasileira de Empresas de Pesquisa. *Critério de Classificação Econômica Brasil*. 2016. Available from: <<http://www.abep.org/criterio-brasil>>. Access on: 10 Jan. 2019.
- AKAMAJIAN, A.; DEMERS, R.; FARMER, A.; HARNISH, R. *Linguistics: An introduction to language and communication*, 5th edition. Cambridge: MIT Press, 2001.
- ALLAN, D. *Oxford Placement Test 1*. Oxford: Oxford University Press, 2004.
- ALVES, U. K. The acquisition of English “-ed” complex codas by Brazilian Portuguese speakers: more than just a matter of constraint demotion. *ROA – 170*, 2004. Available from: <<http://roa.rutgers.edu/files/701-1204/701-ALVES-o-o.PDF>>. Access on: 14 Apr. 2019.
- BARBOSA, P. A.; ALBANO, E. C. Illustrations of the IPA: Brazilian Portuguese. *Journal of the International Phonetic Association*, Cambridge, Inglaterra, v. 34, n. 2, 2004. p. 227-232.
- BISOL, L. A sílaba e seus constituintes. In: NEVES, M. H. M. (org.). *Gramática do português falado*. v. 7: Novos Estudos. Campinas: Ed. da UNICAMP, 1999. p. 701-742.
- BIZZOCCHI, A. Processos de formação lexical das línguas românicas e germânicas: uma nova perspectiva teórica. *Domínios de Lingu@gem*, v. 7, n. 1, 2013. p. 9-39.

- CAMARA JR., J. M. (1970). *Estrutura da Língua Portuguesa*. 45 ed. – Petrópolis, Rio de Janeiro: Editora Vozes, 2013.
- CANTONI, M. M. A epêntese no português brasileiro em uma perspectiva multirrepresentacional. *Gragoatá*, n. 38, 2015. p. 231-246.
- CARDOSO, W. The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, v. 27, n. 4, p. 433-465, 2011.
- CEDERGREN, H. J.; SANKOFF, D. Variable rules: performance as a statistical reflection of competence. *Language*, 50:2. 1974. p. 333-355.
- COLLISCHONN, G. Um estudo da epêntese à luz da Teoria da Sílabas de Junko Ito (1986). *Letras de Hoje*, v. 31, n. 2, 1996. p. 149-158.
- \_\_\_\_\_. Epêntese Vocálica e Restrições de Acento no Português do Sul do Brasil. *Signum: Estud. Ling.*, Londrina, n. 7/1, jun. 2004. p. 61-78.
- CSER, A. *Aspects of the Phonology and Morphology of Classical Latin*. Akadémiai doktori értekezés: Budapest, 2016.
- DABBAGH, A.; NOSHADI, M. Crossing metacognitive strategy awareness in listening performance: An emphasis on language proficiency. *Int. J. Appl. Linguist. English Lit.* 3(6), 2014. p. 234-242.
- DALESZYNSKA, A. *Analysing linguistic variation with Rbrul – a step-by-step guide*. Available from: <[http://www.danielezrajohnson.com/daleszynska\\_rbrul.pdf](http://www.danielezrajohnson.com/daleszynska_rbrul.pdf)>. Access on: 06 Dec. 2019.
- ENNES, M. A.; MARCON, F. Das identidades aos processos identitários: repensando conexões entre cultura e poder. *Sociologias*, Porto Alegre, v. 16, n. 35, 2014. p. 274-305. Available from: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1517-45222014000100010&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-45222014000100010&lng=en&nrm=iso)>. Access on: 07 Dec. 2019.
- GOMES, C. A. Para além de pacotes estatísticos VARBRUL/GOLDVARB e RBRUL: qual a concepção de gramática? *Revista do GELNE*, v. 14, número especial, 2012. p. 259-272.
- GREY, A.; PILLER, I. Sociolinguistic ethnographies of globalisation. In: TUSTING, K. (Ed.). *The Routledge handbook of linguistic ethnography*. Vol. Routledge handbooks in applied linguistics. London: Routledge, Taylor and Francis Group, 2019. p. 54-69.
- GUY, G. R.; ZILLES, A. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola Editorial, 2007.

- HALL, J. K. *Teaching and researching: Language and culture*. Harlow: Longman, 2002.
- HORVATH, B. M.; HORVATH, R. J. A closer look at the constraint hierarchy: Order, contrast and geographic scale. *Language, Variation and Change*. 15, 2003. p. 143-170.
- HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. 2nd ed. New York: Wiley, 2000.
- JOHN, P.; CARDOSO, W. On syllable structure and phonological variation: The case of i-epenthesis by Brazilian Portuguese learners of English. *Ilha Desterro*, Florianópolis, v. 70, n. 3, p. 169-184, Dec. 2017. Available from: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S2175-80262017000300169&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-80262017000300169&lng=en&nrm=iso)>. Access on: 24 Nov. 2019.
- JOHNSON, D. E. *Rbrul version 2.29: A variable rule application in R*. 2015.
- KENT, R. D.; READ, C. *Análise Acústica da Fala*. Tradução de Alexsandro Meireles. São Paulo: Cortez, 2015.
- LOPES, M. O. As “novas” faces do imperialismo norte-americano: balanço teórico. In: Anais do Simpósio Lutas Sociais na América Latina, 2, Crise das democracias latino-americanas: dilemas e contradições. *Anais...* Gepal, Londrina: UEL, 2006 Available from: <<http://www.uel.br/grupo-pesquisa/gepal/segundosimposio/marianadeoliveiralopes.pdf>>. Access on: 09 Feb. 2019.
- LUCENA, R. M.; ALVES, U. K. Implicações dialetais (dialeto gaúcho vs. paraibano) na aquisição de obstruintes em coda por aprendizes de inglês (L2): uma análise variacionista. *Revista Letras de Hoje*, Porto Alegre (RS). v. 45, n. 1, Jan.-Mar., 2010.
- MADEIRA, A. “Novas narrativas sobre os Estados Unidos” (resenha do livro *Americanos: representações da identidade nacional no Brasil e nos EUA*, de Lúcia Lippi Oliveira, Belo Horizonte, Editora da UFMG). *Revista Sociedade e Estado*, XV (2): jun.- dez., Rio de Janeiro: 2002. p. 388-398.
- MASSINI-CAGLIARI, G. *A música da fala dos trovadores: desvendando a prosódia medieval*. São Paulo: Ed. da Unesp, 2015.
- MONARETTO, V. N. O. Sequência de Obstruintes no Interior de Palavras: retrato em duas modalidades. *Estudos Linguísticos e Literários*, v. 1, 2015. p. 205-230.

- \_\_\_\_\_.; QUEDNAU, L. R.; HORA, D. As consoantes do português. In: BISOL, L. (Org.). *Introdução a estudos de fonologia do português brasileiro*. 5 ed. Porto Alegre: EDIPUCRS, 2014, v. 1. p. 202-235.
- PIKE, K. L.; FRIES, C. C. Coexistent phonemic systems. In: BREND, R. (Ed.). *Kenneth L. Pike: Selected writings to commemorate the 60th birthday of Kenneth Lee Pike*. The Hague: Mouton, 1972. p. 51-73.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. 2018. Available from: <<https://www.R-project.org>>. Access on: 10 Dec. 2018.
- REIS, F. S.; LUCENA, R. M. Variabilidade na produção das oclusivas coronais entre consoantes heterossilábicas por aprendizes campinenses de inglês como L2. *Domínios de Linguagem*, v. 13, n. 4, 2019. p. 1596-1635.
- ROY, J.; OETTING, J. B.; MOLAND, C. W. Linguistic constraints on children's overt marking of BE by dialect and age. *Journal of Speech, Language, and Hearing Research*, 56, 2013. p. 933-944.
- SANKOFF, D.; TAGLIAMONTE, S.; SMITH, E. *GoldVarb X: a variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto, 2005.
- SCHNEIDER, A.; SCHWINDT, L. C. S. A epêntese vocálica medial em PB e na aquisição de inglês como LE: uma análise morfofonológica. *Letras de Hoje*, Porto Alegre, v. 45, n.1, jan./mar. 2010. p. 16-26.
- SEIDLHOFER, B. *Habeas corpus and divide et impera: 'Global English' and applied linguistics*. In: MILLER, K. S.; THOMPSON, P. (Eds.). *Unity and diversity in language use*. London: Continuum, 2002. p. 198-217.
- SILVA, J. C. Cinema, identidade cultural latino-americana e o imperialismo estadunidense. Trabalho de conclusão de curso. ECA/USP, São Paulo, 2015.
- SILVEIRA, R. Pronunciation instruction and syllabic-pattern discrimination. *DELTA*, São Paulo, v. 27, n. 1, 2011. p. 21-36. Available from: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=Soi02-44502011000100002&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=Soi02-44502011000100002&lng=pt&nrm=iso)>. Access on: 19 May 2019.
- TARALLO, F. *A pesquisa sociolinguística*. 8ª ed. São Paulo: Ática, 2007.



XAVIER, M. F.; MATEUS, M. H. M. (Org.). *Dicionário de termos linguísticos*. v. 1. Lisboa: Cosmos, 1990.

**Recebido em 27/05/2019.**

**Aprovado em 28/12/2019.**