

RHOTICS IN FINAL SYLLABLE CODA OF INFINITIVE VERB FORMS IN “CAIPIRA” PORTUGUESE SPOKEN IN SÃO PAULO STATE

Selmo Ribeiro FIGUEIREDO JUNIOR*

- **ABSTRACT:** Within the framework of Pluridimensional Dialectology (RADTKE; THUN, 1996; THUN, 2000, 2005 etc.), this study presents both a record and an analysis of variants in final syllable coda of the infinitive ending followed by a break, as they are currently spoken in the Médio Tietê region, the birthplace of São Paulo state’s “Caipira” culture. The data were collected by Figueiredo Jr. (2019) in ten localities in the region: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari, and Piracicaba. The methodological instrument applied is the Phonetic-Phonological Questionnaire made by “Atlas Linguístico do Brasil” (COMITÊ NACIONAL DO PROJETO ALiB, 2001). Eighty informants partook in the research, evenly women and men, young (18–36) and old (≥ 55) individuals, and with low and high education. The variants collected are: retroflex approximant ([ɻ]), alveolar tap ([ɾ]), phonetic zero ([Ø]), and alveolar trill ([r]). The analysis is based on descriptive statistics and pluridimensional dialectological cartography, from which a set of relevant conclusions is drawn. One of them is that the retroflex approximant occurs as the most frequent variant from a broad perspective and, from a correlational viewpoint, another one is that, according to evidence, the external factor influencing the prevalence of the retroflex approximant is not diastratic, nor diasexual, nor diatopic, but rather diagenational.
- **KEYWORDS:** pluridimensional dialectology; geolinguistics; phonetics; São Paulo state’s Portuguese; Caipira dialect; Médio Tietê; archiphoneme /R/ in syllable coda.

Introduction¹

A set of oral corpora was collected from 2016 to 2017 through the methodology of Pluridimensional Dialectology (RADTKE; THUN, 1996; THUN, 2000, 2005, etc.) in

* Charles University (CUNI) – Czech Republic. Assistant Professor in the Department of Romance Studies. selmo.figueiredo@ff.cuni.cz. ORCID: 0000-0001-8367-0306.

¹ I acknowledge the critical reading of some colleagues of mine. They are Felício W. Margotti (Federal University of Santa Catarina, UFSC), Greize A. da Silva (Federal University of Tocantins, UFT), Valter P. Romano (UFSC), Luiz Egito de S. Barros (Federal University of Piauí), Manoel M. Santiago-Almeida (University of São Paulo, USP), Juscelino F. do Nascimento (UFPI), and Mario Ruiz Moreno (Johannes Gutenberg-Universität Mainz, JGU). I also thank this journal’s anonymous reviewers for their suggestions.

order to mainly draw up a linguistic (i.e. phonetic-phonological and semantic-lexical) atlas—published as a doctoral dissertation (FIGUEIREDO JUNIOR, 2019)—of the so-called “Caipira” dialect spoken in the inland region of Médio Tietê, São Paulo state, Brazil. This region is considered the center of the “Caipira” culture, where 80 informants were interviewed in ten of its oldest localities. After a short introduction to rhotics in general, this paper focuses on both presenting and analyzing the data concerning the covariants competing to fill the variable /R/—as an archiphoneme in the final syllable coda of the infinitive ending followed by a break—in perspective with the diatopic, diastratic, diasexual, and diagenational dimensions. By doing so, this study contributes to the dialectological understanding of the “Caipira” dialect spoken in São Paulo state whose precursor efforts go back to Amadeu Amaral (1920).

Rhotics—which are phones associated with both graphemes <r> and <rr> in Portuguese—are a subject matter of a long phonetic and phonological debate involving countless natural languages. Many scholars even argue that determining a class of rhotics is an arbitrary act (cf., e.g., SCOBIE, 2006; WIESE, 2011; SEBREGTS, 2015; CHARBOT, 2019). This argument is based on the premise that a phone cannot be a priori predicted from a phoneme. In other words, a phone can only be undoubtedly determined through its a posteriori phonetic realization.

Charlot (2019) puts forward a definition of phonological rhotics (not phonetic rhotics) based on their phonotactic behavior rather than their phonological representations. For this purpose, he states that rhotics tend to exhibit both procedural and diachronic stability. The former concerns these segments (implicated in phonological processes) as being able to phonetically vary in an arbitrary way without perturbing the phonological processes themselves. The latter involves these segments as being able to diachronically vary without causing a realignment in a phonological system. In two parts, his definition follows.

1. A rhotic is a segment which may occupy specific syllabic positions—that of the secondary element in branching onsets or codas—and functions as a sonorant regardless of its phonetic instantiation.
2. A rhotic demonstrates PROCEDURAL and DIACHRONIC STABILITY: its phonotactic status as a sonorant does not change even when the rhotic is subject to variation due to either diachronic evolution or synchronic processing—for example even if the rhotic is realized as an obstruent. (CHARBOT, 2019, p.11, small capitals in the original).

This definition is unduly generic due to its assumption that does not imply phonetic substance, but it avoids some problems in the literature on rhotics, such as attempts at establishing univocity between phonological and phonetic objects in natural languages. This matter is especially dealt with in the case of rhotics, for they appear to be among the segments most subject to phonetic variation in natural languages.

Barry (1997) on this variation, for example, tackles the argument that all instantiations of /R/ ultimately stem from a universal non-lateral liquid as well as the hypothesis that the alveolar trill has primacy. In this regard, many authors address sets of rhotics (e.g., WIESE, 2011; LADEFOGED; MADDIESON, 1996; LINDAU, 1985). Relevant criticism of proposals for these sets emerges in favor of the effort to phonologically classify rhotics instead of attempts based on diachronic relations or on orthographic, symbolic conventions. A non-phonological effort results in a conventional, arbitrary classification. Charbot's (2019) phonological attempt is to be considered and perhaps revised to be even leaner.

In any event, such productive variability of rhotics in countless languages is a feature to be highlighted (cf., e.g., COHEN, LAKS, SAVU, 2019: on Modern Hebrew; SEBREGTS, 2015: on Dutch; ROMANO, 2013: on Italian varieties; PATIN, 2013: on Washili, a variety of Shingazidja²; SCHILLER, 1998: on German; among so many other languages).

Portuguese, in particular, is especially productive regarding the allophonic variation of /R/. For instance, Comiotto and Margotti (2019) investigate rhotic variants in Portuguese spoken in Rio Grande do Sul and in Santa Catarina (both Brazilian states) in contact with Italian varieties. Brandão *et al.* (2017) focus on rhotic variants in prevocalic and postvocalic positions in the urban variety of Portuguese spoken in São Tomé (capital of São Tomé and Príncipe). Callou and Serra (2012) approach the deletion of /R/ in the final syllable coda in Portuguese spoken in Salvador and in Rio de Janeiro (Brazilian cities).

After this short introduction, the terms of the research brought in this work are presented next.

Locus and methodology

This section presents the network of localities visited and the methodology applied in Figueiredo Jr.'s (2019) research, which is discussed and analyzed afterwards. This research is focused on the phonetic variation associated with the archiphoneme /R/ in the final syllable coda of the infinitive ending followed by a break in perspective with the diatopic, diastratic, diasexual, and diagenational dimensions.

The data were collected in ten cities in São Paulo state that lie in the Médio Tietê region, the cradle of the “Caipira” culture. These cities are: Santana de Parnaíba (originated in 1561), Pirapora do Bom Jesus (1725), Araçariguama (1590), São Roque (±1665), Sorocaba (1654), Itu (1610), Porto Feliz (1721), Tietê (±1500), Capivari (±1760), and Piracicaba (1766). The variety of Portuguese spoken there represents one of the earliest phases of the lusitanization of Brazil beginning in São Vicente city

² Shingazidja is a Bantu language spoken on Grande Comore, an island that belongs to Comoros, an archipelago off Africa's east coast.

(São Paulo nowadays) as of 1532. Due to direct influence from a series of passages—carried out by manifold-purpose missions called “*bandeiras*” (on land) and “*moçôes*” (by rivers) as well as by the so-called “*tropeiros*” (‘drovers’) between the 16th and 18th centuries—Médio Tietê incorporated cultural (and linguistic) characteristics in its own way (cf. CÂNDIDO, 2001).

Within the theoretical and methodological framework of Pluridimensional Dialectology (RADTKE; THUN, 1996; THUN, 2000, 2005 etc.), interviews were conducted with 80 informants. Each locality provided eight informants, four men and four women. Furthermore, there were four criteria that all informants had to meet:

- a. to belong to one of two age groups: either to GI (18–36) or GII (55 and upwards);
- b. to belong to one of two social/school classes: either to Cb (ranging from illiteracy to high school incomplete) or Ca (from higher education at least begun);
- c. to have lived in the respective locality for at least three quarters of life;
- d. to have lived in the respective locality for at least the last five years uninterruptedly.

These criteria fulfilled generate an even distribution of informants. Four profiles are represented for each of both sex groups: CaGII, CbGII, CaGI, and CbGI. Thus, each of these profiles expresses a data crossing between both the diastratic (Cx) and diagenational (Gy) variables/dimensions, synthesized as CxGy. Since the diasexual variable (Wz, whose concrete value can be either Wf for female individuals or Wm for male ones) was also methodologically controlled, other two data crossings are systematically possible: CxWz e WzGy.

The methodological instrument applied is the Phonetic-Phonological Questionnaire by *Atlas Linguístico do Brasil* (COMITÊ NACIONAL DO PROJETO ALiB, 2001) with 159 inductive questions aiming at providing conditions to observe the behavior of a set of phenomena, among which rhotics in syllable coda, in simple onset, and in C2 position of complex onset.

This paper’s specific object is the archiphoneme /R/ in the final syllable coda of the infinitive ending followed by a break. This variable is subsequently referred to as E1, following Figueiredo Jr. (2019, *passim*). Table 1 ahead shows the data collected, produced by informants answering questions 18, 36, and 153. These questions induced them to enunciate the “pretext-words”³ <varrer> (‘to sweep’), <botar> (‘to lay [eggs]’) e <sair> (‘to go out, to leave’), respectively, and in Figueiredo Jr.’s (2019) research were posed by the inquirer as follows (Q = question):

³ I am referring to these words in this way because they are not meant to be analyzed as a whole, but rather as a pretext to make informants pronounce the relevant segments (in **bold**) under study.

- Q18 *Para limpar o chão, o que é preciso fazer com este objeto que a pessoa está segurando?*
 ‘What is necessary to do with this object this person [in a picture] is holding in order to clean the floor?’
- Q36 *Quando a galinha sai cantando do ninho é sinal de que ela provavelmente acabou de _____ ovo.*
 ‘When a hen gets out of its nest singing, it is a sign that it has probably just [laid] an egg.’
- Q153 *Qual é o contrário/oposto de entrar?*
 ‘What is the opposite of entering[/going into]?’

In the case of Q18, its enunciation occurred as a picture of a person sweeping the floor with a broom was simultaneously shown. In the case of Q36, <botar> was expected as the answer to fill in the blank.

Data and initial analysis

Table 1 shows the data. They are initially discussed and analyzed from a broad perspective, i.e. without being distributed across the profiles controlled in the study (such as CaGII, CaGI, CbGII e CbGI).

Table 1 – Data for E1 with no crossings with extralinguistic variables

PPQ questions	Variants collected and their valid values registered				Σ	
	[ɹ]	[r]	[Ø]	[r]		
Q18 <varrer>	36	6	37	1	80	
Q36 <botar>	26	10	44	0	80	
Q153 <sair>	54	13	10	3	80	
Σ	absolute value	116	29	91	4	240
	relative value	48%	12%	38%	2%	100%

Q = question • PPQ = Phonetic-Phonological Questionnaire

Source: Figueiredo Jr. (2019, p.180).

The variants registered are four: 1. retroflex approximant, [ɹ]; 2. alveolar tap, [r]; 3. phonetic zero, [Ø]; 4. alveolar trill, [r]. As seen in Table 1, the phonetic zero as a variant associated with Q153 has a frequency expressively lower than those the other

variants have. In this regard, I suggest the hypothesis (to be tested in future studies) that [sa'i] does not represent a significant occurrence for the infinitive ending at issue because such a phonetic realization is statistically preferred for <saí> ('[I] went out'), conjugated in the indicative mood, “*pretérito perfeito simples*” ('simple perfect past tense'), first person singular of <sair>.

The most frequent rhotic variant for E1 from a broad perspective is [ɹ]. The retroflex approximant in this variation stands out in direct competition with the phonetic zero. This competition is particularly fierce on account of the pretext-word <varrerɹ>.

We can look for what Figueiredo Jr. (2019) calls absolute phonetic-phonological norm. It refers to a variant that from a broad perspective (i.e. with no correlations with extralinguistic variables) has both a regular diatopic distribution and a relative frequency (RF) greater than 50% among the other valid variants. A variant in regular diatopic distribution is understood as one occurring throughout the network of localities.

From a non-correlational perspective and considering all the data, no such a norm was registered, but [ɹ] has almost gotten there. This variant is the best candidate for the post. It has regular diatopic distribution⁴ and 48% RF.

On the one hand, considering only the data related to Q153, the retroflex variant would be an absolute norm, since this variant has both regular diatopic distribution and 68% RF. On the other hand, considering only the data related to Q36, the phonetic zero would be the one to be an absolute norm, with regular diatopic distribution and 55% RF.

By distributing the empirical results among the informant groups, we can better comprehend the dialectological behavior of E1. Table 2 displays a distribution involving both the diastratic and diagenational dimensions. The former refers to high and low social/school classes, Ca and Cb. The latter relates to the older and younger age groups, GII and GI. This crossing is synthesized as CxGy.

Table 2 – Data for E1 in crossing CxGy

Var.	CaGII		CaGI		CbGII		CbGI	
	AF	RF	AF	RF	AF	RF	AF	RF
[ɹ]	27	45%	34	57%	18	30%	37	62%
[r]	15	25%	8	13%	4	7%	2	3%
[Ø]	17	28%	18	30%	35	58%	21	35%
[r]	1	2%	0	0%	3	5%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Var. = variant • AF = absolute frequency • RF = relative frequency
 Ca/b = high/low school class • GII/I = older/younger age group

Source: Figueiredo Jr. (2019, p.181).

⁴ All mentions of diatopic information in this work can be cartographically seen in the dialectological charts section.

Given that a relative norm is a notion that coincides with absolute norm (as defined previously) except for the difference that it is so only in a specific profile or informant group instead of being so in general (cf. FIGUEIREDO JUNIOR, 2019), Table 2—and, as a cartographical equivalent to this table, Figure 1 in the next section—allows us to see that the variant [ɹ] is a relative norm among younger individuals, not only among those of low class (62% RF), but also among those of high class (57% RF). This provides both empirical and quantitative evidence for the prevalence of the retroflex approximant among younger speakers nowadays. It goes against the popular expectation that this phone would be more productive among older speakers, mostly among those belonging to low class.

In fact, the data point towards the prevalence of [ɹ] among high-class older individuals (45% RF), not among those of low class (30% RF). As to young individuals, however, the popular expectation that those of low class would produce the retroflex approximant in a more productive way is supported by empirical results, even though the difference of 5% is not significant.

Another finding concerns the phonetic zero. It has a significant rate of 58% RF among low-class older speakers and has regular diatopic distribution. This adds up to the variant [Ø] being a relative norm.

As for the alveolar tap, if the sequence ‘CaGII, CaGI, CbGII, and CbGI’ is regarded as a decreasing one, the values reached by the phone [r] nearly represent an increasing geometric progression in relation to the mentioned sequence with approximate ratio two ($q \cong 2$). This outcome implies a correlation according to which the younger people are and the less schooling they have, the lower the relative frequency of the alveolar tap is among them.

When it comes to a CxWz data crossing—i.e. a crossing with both diastratic (Ca and Cb) and diasexual (Wf and Wm, standing for feminine and masculine informants, respectively) dimensions—Table 3 organizes it.

Table 3 – Data for E1 in crossing CxWz

Var.	CaWf		CaWm		CbWf		CbWm	
	FA	FR	FA	FR	FA	FR	FA	FR
[ɹ]	31	52%	30	50%	23	38%	32	53%
[r]	12	20%	11	18%	4	7%	2	3%
[Ø]	17	28%	18	30%	30	50%	26	43%
[r]	0	0%	1	2%	3	5%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Var. = variant • AF = absolute frequency • RF = relative frequency
 Ca/b = high/low school class • Wf/m = feminine/masculine sex

Source: Figueiredo Jr. (2019, p.182).

We can infer from Table 3 that the retroflex approximant among high-class feminine individuals almost constitutes a relative norm with its 52% RF, ergo above the minimum of 50%. What keeps it from having such status is only the fact that the variant at issue is not observed throughout the network of localities. In other words, [ɹ] in CaWf does not have regular diatopic distribution, what is seen as imagery in the dialectological charts section. Yet, this variant does have the status of relative norm among low-class masculine individuals, with 53% RF.

Moreover, there is a competition between the alveolar tap and phonetic zero for filling E1 that puts both social/school classes on opposite sides from a quantitative perspective. While both high-class groups, CaWf and CaWm, enunciate [r] with 20% and 18% RF, respectively, both low-class groups, CbWf and CbWm, articulate the same phone with 7% and 3% RF, respectively. Apropos [Ø], both high-class groups, CaWf and CaWm, produce this variant with 28% and 30% RF, respectively, while both low-class groups, CbWf and CbWm, produce the same phone with 50% and 43% RF, respectively.

In reference to E1 interacting with both the diasexual and diagenational variables/dimensions, Table 4 frames the relevant data.

Table 4 – Data for E1 in crossing WzGy

Var.	WfGII		WfGI		WmGII		WmGI	
	FA	FR	FA	FR	FA	FR	FA	FR
[ɹ]	18	30%	36	60%	27	45%	35	58%
[r]	10	17%	6	10%	9	15%	4	7%
[Ø]	29	48%	18	30%	23	38%	21	35%
[r]	3	5%	0	0%	1	2%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Var. = variant • AF = absolute frequency • RF = relative frequency
Wf/m = feminine/masculine sex • GII/I = old/young age group

Source: Figueiredo Jr. (2019, p.183).

The retroflex approximant is a relative norm among both young groups. In this crossing, [ɹ] has regular diatopic distribution and relative frequencies of 60% among feminine young individuals and of 58% among masculine young individuals.

Among feminine old individuals, the phonetic zero was close to being a relative norm with its 48% RF and presence throughout the localities visited. However, even without being a relative norm, [Ø] overcomes [ɹ] in the same crossing by having 18 percentage points more. This fact is even more noteworthy when one knows that the retroflex approximant is the most productive variant from a broad perspective (s. Table 1).

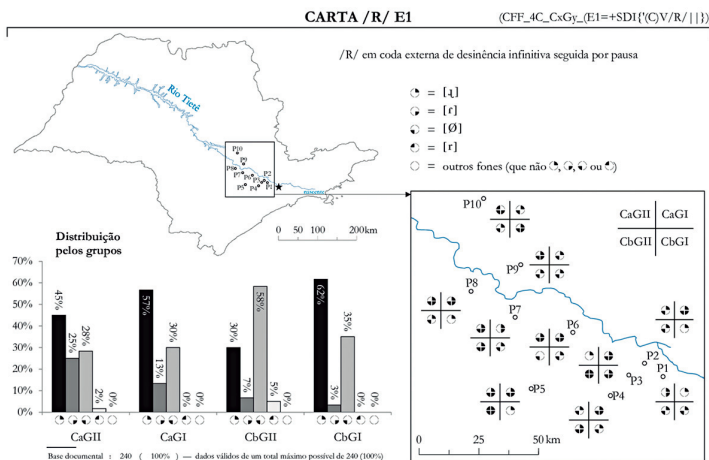
Here, we finish an analysis based on Tables 1, 2, 3, and 4. Essentially, they display data crossings that help us see correlations between linguistic data on the one hand and extralinguistic data on the other hand. The linguistic data refer to variants that compete to fill the variable E1, i.e. archiphoneme /R/ in the final syllable coda of the infinitive ending followed by a break. The extralinguistic data concern the diastatic, diasexual, and diagenational variables/dimensions. In the next section, we analyze the data based on dialectological charts.

Dialectological charts and complementary analysis

The previous tables make it possible for us not only to cast a clean glance at the data and to consequently have a first approach to the results, but also to use them as a documentary basis for us to draw up linguistic charts (or rather dialectological charts, to be more accurate) in regard to E1 as the phonetic-phonological variable at issue. For this reason, this paper is specifically subsumed under the studies of geolinguistics, which is here understood as a branch of dialectology that is particularly interested in mapping dialectal material.

Such charts were made through the methodology of Pluridimensional Dialectology (RADTKE; THUN, 1996; THUN, 2005, 2000). Its terms and symbols are introduced ahead, so that anyone not familiarized with pluridimensional geolinguistics can navigate what is necessary to correctly read and interpret its charts. To achieve this purpose, let us start with Figure 1. It correlates E1 and extralinguistic variables, namely diatopic, diastratic (Cx), and diagenational (Gy) ones. These extralinguistic variables are also known as dimensions within pluridimensional geolinguistics.

Figure 1 – Chart for E1 in crossing CxGy



Source: Figueiredo Jr. (2019, p.1045).

At the top, we see the chart’s simplified title in the center and its technical title on the right. The latter can be ignored in view of the relevant purposes in this work. A little below, the linguistic variable under consideration—‘/R/ in the final syllable coda of the infinitive ending followed by a break’—is announced. The map at the top left represents São Paulo state’s political-administrative limits, the Tietê river, the localities (or points) in the Médio Tietê region under study, and the capital of the state as a star. At its right side, we see the variants registered and the symbols assigned to them. These symbols help us interpret the map enlarged at the bottom right, at whose top right it is found the biggest cross that serves as the key to the cartographical interpretation. The biggest cross comprises the group captions CaGII, CaGI, CbGII and CbGI, which were already introduced previously. The ten smaller crosses placed near their respective localities/points are interpreted in concordance with the mentioned key. The ten points are as follows:

Table 5 – Network of localities/points

Point	Name	Emergence	Population*	Point	Name	Emergence	Population*
P1	Sant. de Parnaíba	1561	139.447	P6	Itu	1610	173.939
P2	Pir. do Bom Jesus	1725	18.895	P7	Porto Feliz	1721	53.098
P3	Araçariguama	1590	22.364	P8	Tietê	±1500	42.076
P4	São Roque	±1665	91.016	P9	Capivari	±1760	55.768
P5	Sorocaba	1654	679.378	P10	Piracicaba	1766	404.142

* Estimated population in 2019 (IBGE, 2021)

Source: Author’s elaboration.

Figure 1 contains a statistical graph (at the bottom left) that distributes the valid data among the four informant groups. The percentage values are relative frequencies of the phonetic variants produced by the groups. Lastly, a footnote reports ‘documentary basis: 240 (100%) valid pieces of data out of a possible total of 240 (100%)’ for the chart.

As seen in Figure 1, there is no variation (i.e. there is only a variant) in a quadrant of each of six localities. These localities are: Santana de Parnaíba (P1), Pirapora do Bom Jesus (P2), Araçariguama (P3), Sorocaba (P5), Itu (P6), and Tietê (P8). It is the phonetic zero enunciated by CbGII in P6 and the retroflex approximant enunciated predominantly by CbGI in the other points (P2, 5, and 8).

When it comes to the variation formed by a competition specifically between two variants in the quadrants of Figure 1, there are two combinations: [ɹ]-[r] and [ɹ]-[Ø]. The former occurs twice (P1 and 7) on account of high class. Namely CaGII in P1 and CaGI in P7. The latter is more prevalent and diffusely appears 14 times in the diatopic dimension. A fact that stands out is that 10 out of these 14 occurrences on the chart are on account of low class.

There is also the variation formed by a competition among three variants in two combinations: [r]-[Ø]-[r] and [ɹ]-[r]-[Ø]. The former is only documented in Piracicaba (P10) among CaGII members. The latter with 15 occurrences—12 out of which produced by high class—happens throughout the network of localities not counting Santana de Parnaíba (P1).

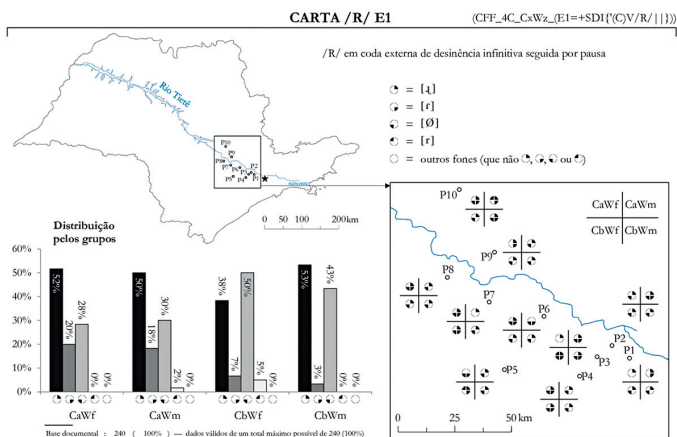
The variation created by a competition among all four variants that are objects of our cartography is solely noticed in Araçariçuama (P3) and Sorocaba (P5) articulated by low-class old individuals.

Figure 1 also reveals a prevalence of [ɹ] among both young groups as well as among high-class old speakers. The phonetic zero comes as the second most productive variant from a broad perspective and as the first most productive variant among low-class individuals with 58% RF against 30% RF obtained by the retroflex approximant. Taking into account all four informant groups, CbGI is the one whose members say [ɹ] the most and [r] the least.

Remarkably, Figure 1 does not record the alveolar trill among young speakers. Another detail that deserves our attention is an inverted symmetry that takes place between CaGI and CbII, which are entirely opposed to each other apropos their respective profiles. This inverted symmetry is caused by the fact that CaGI pronounces [ɹ] with 57% RF and [Ø] with 30% RF, almost in the same ratio that CbGII pronounces [Ø] with 58% RF and [ɹ] with 30% RF. Hence, it refers to a possible change.

Interpretations and inferences based on Figure 1 do not peter out here with those performed above. They are just some that seem more protruding. Next, the same analytical exercise is done, grounded in Figure 2, which makes explicit a correlation between E1 (/R/ in the final syllable coda of the infinitive ending followed by a break) and the diatopic, diastatic (Cx), and diagenational (Wz) dimensions.

Figure 2 – Chart for E1 in crossing CxWz



Source: Figueiredo Jr. (2019, p.1046).

Bearing in mind only both upper quadrants of Santana de Parnaíba (P1), Araçariçuama (P3), and Porto Feliz (P7) in Figure 2, they present a quadrant (an informant group) each in which there is no variation going on, i.e. only one variant occurs. It is the retroflex approximant spoken by CaWf in both P1 and P3 as well as by CaWm in P7. Therefore, by high-class individuals for that matter.

As to the variation formed by two variants, there are three combinations: [ɹ]-[r], [r]-[Ø], and [ɹ]-[Ø]. The former two happen twice each, and the latter does 19 times. Though both occurrences of the first combination on the chart are expressed by high class informants, these occurrences are far apart from one another: one in Santana de Parnaíba (P1) and another in Capivari (P9). Concerning both instances of the second combination, they come from high class too and this time they are close to one another: one in Sorocaba (P5) and another in Itu (P6). Lastly, the third combination is observed throughout the network of localities aside from Araçariçuama (P3). It instantiates at least once in each of the possible quadrants. Out of its 19 occurrences, 14 originate from low class. In view of all high-class women, this combination is found only once, in Pirapora do Bom Jesus (P2).

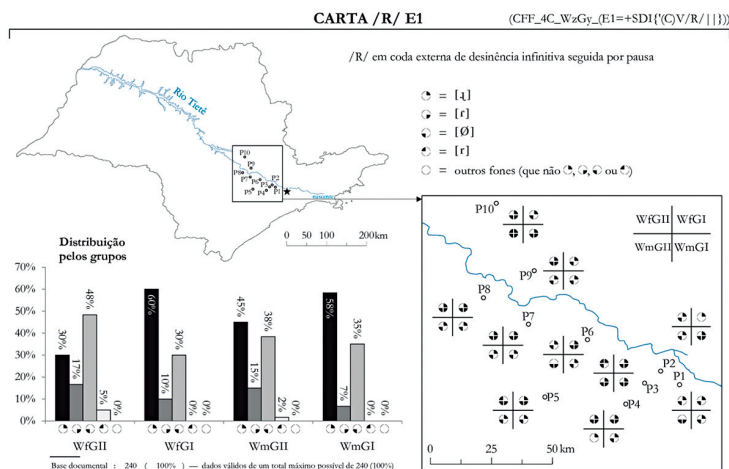
In respect to the variation formed by three variants in competition, there are two combinations: [ɹ]-[r]-[r] and [ɹ]-[r]-[Ø]. The former has two instances present in Figure 2 and is spotted in both Sorocaba (P5) and Piracicaba (P10), which are far apart from one another and whose informant groups are diastatically and diagenetically different. The latter is more common, cartographically occurring 11 times, five of which take place among high-class women, in five localities: São Roque (P4), Itu (P6), Porto Feliz (P7), Tietê (P8), and Piracicaba (P10). This second combination occurs in the diatopic dimension in a more consistent way in the upper part—having the Tietê river as the spatial reference—of the Médio Tietê region as framed in the cartographic representation. On the subject of the variation formed by four variants in competition, there is solely one, instantiating in Araçariçuama (P3) among low-class women.

Figure 2 quantitatively implies a probable pattern present in both high-class groups concerning all variants. To demonstrate this, let us first consider the variants that have more significant results, viz. the retroflex approximant, alveolar tap, and phonetic zero. These three phones have 52%, 20%, and 28% RF, respectively, among feminine individuals as well as 50%, 18%, and 30% RF, respectively, among masculine individuals. Note that this parallelism between both feminine and masculine groups exists within a two-percentage-point margin throughout the variational spectrum shaped by the variants at issue.

Moreover, the retroflex approximant constitutes the most frequent variant among all speaker groups barring low-class feminine group, in which the phonetic zero is the most frequent variant, with its 50% RF against 38% RF of [ɹ]. In passing, the phonetic zero is produced by individuals from both low-class groups with percentage values considerably higher than those found in both high-class groups. The opposite is the case that involves the alveolar flap. Its values are comparatively high in both high classes (CaWf: 20%; CaWm: 18% RF) and low in both low classes (CbWf: 7%; CbWm: 3%).

The remarks above put the linguistic data associated with E1 (archiphoneme /R/ in final syllable coda of the infinitive ending followed by a break) in perspective with both diastatic and diasexual extralinguistic variables. Hereinafter, it is exhibited the interaction between the same phonetic-phonological variable (E1) on the one hand and the diatopic, diasexual, and diagenational variables on the other hand starting with Figure 3.

Figure 3 – Chart for E1 in crossing WzGy



Source: Figueiredo Jr. (2019, p.1047).

Firstly, one can notice based on Figure 3 that there is no variation among both young women in Pirapora do Bom Jesus (P2) and old men in São Roque (P4). In this case, it is about the retroflex approximant only.

On the issue of the variation forged by two variants, there are two combinations: [r]-[Ø] and [ɹ]-[Ø]. The former solely circulates among old individuals with two occurrences on the chart, the one in Santana de Parnaíba (P1) and the other in Itu (P6). The latter is distributed throughout the network of localities excluding Araçariçuama (P3) with 18 instances of variable variants occupying different quadrants among the four possible. Twelve of such instances are assigned to both masculine groups throughout the network of localities with the exception of Piracicaba (P10) besides P3.

As regards the variation that consists of three variants in competition, there are three combinations: [ɹ]-[Ø]-[r], [r]-[Ø]-[r], and [ɹ]-[r]-[Ø]. The former two are insignificant. Each of them has only one occurrence on the chart under consideration (Figure 3), viz. in Araçariçuama (P3) and in Sorocaba (P5), respectively, both of them produced by old women. The latter is spread in all points but for Santana de Parnaíba (P1) with 15 occurrences, 11 of which coming from both feminine groups in all localities except for Pirapora do Bom Jesus (P2) besides P1.

Regarding the maximum variation that is possible within the cartographical representation model adopted in this work (to wit, among four variants), there is only one instance, in young men's speech in Piracicaba (P10).

Furthermore, Figure 3 enables us to realize that the results suggest that there is a pattern in three informant groups involving three variants. First, WfGI enunciates [ɹ] with 60%, [Ø] with 30%, and [r] with 10% RF. Second, WmGII articulates the same variants with 45%, 38%, and 15%, respectively. Third, WmGI speaks with 58%, 35%, and 7% RF, respectively. That is, a descending line lies in the groups at issue when all the mentioned values are kept in view. This in addition to the fact that group WfGII uses [Ø] as the most frequent variant—with 48% against 30% RF of the most frequent variant, retroflex approximant, in the other three groups—leads us to the demonstration that old women are the ones (compared to the other groups in our research sample in Médio Tietê) who most apply the phonetic zero to fill the variable /R/ in final syllable coda of the infinitive ending followed by a break.

Conclusions

On the one hand, the analysis developed on the basis of the statistical tables provides a zoom-out approach to the results through which it is possible to verify what is most typical in the field research under consideration. On the other hand, the analysis performed on the basis of the dialectological charts signifies a zoom-in approach that offers a visualization of what rarely occurs. This last viewpoint is what differentiates Pluridimensional Dialectology from other disciplines that deal with linguistic variation. This said, we finish this paper with some conclusions below.

Firstly, E1 (archiphoneme /R/ in final syllable coda of the infinitive ending followed by a break) from a broad perspective is a variable mostly filled by the retroflex approximant (48% FR), by the phonetic zero (38% FR), by the alveolar tap (12% FR), and lastly by the alveolar trill (2% FR).

Secondly, a correlational perspective between E1 and both diastratic and diagenational variables allows us to see that:

1. the retroflex approximant is a relative norm only among young individuals, from both low class (62% RF) and high class (57% RF);
2. the phonetic zero is a relative norm (58% de RF) solely among low-class old speakers;
3. the younger people are and the less schooling they have, the lower the relative frequency of the alveolar tap is among them—CaGII: 25%; CaGI: 13%; CbGII: 7%; and CbGI: 3% RF—whose values nearly constitutes an increasing geometric progression with approximate ration two ($q \cong 2$);
4. the alveolar trill is not found in young informants' speech; and

5. there is an inverted symmetry between CaGI and CbII—which are entirely opposed to each other apropos their respective profiles—caused by the fact that CaGI pronounces [ɹ] with 57% RF and [Ø] with 30% RF, almost in the same ratio that CbGII pronounces [Ø] with 58% RF and [ɹ] with 30% RF. Hence, it refers to a possible change in progress.

Thirdly, a correlational perspective between E1 and both diastatic and diasexual variables evinces that:

6. the retroflex approximant is almost a relative norm among high-class feminine individuals. It meets the quantitative criterion, with 52% FR (two percentage points greater than the minimum value, 50%), but it does not fulfill the diatopic criterion (having regular diatopic distribution) by not being present in one of the localities, Sorocaba (P5);
7. the retroflex approximant is a relative norm in low-class masculine group (53% RF);
8. the variation between [ɹ] and [Ø] puts both social/school classes on opposed sides: while both high-class groups, CaWf and CaWm, enunciate [ɹ] with 20% and 18% RF, respectively, both low-class groups, CbWf and CbWm, articulate the same phone with 7% and 3% RF, respectively. Apropos [Ø], both high-class groups, CaWf and CaWm, produce this variant with 28% and 30% RF, respectively, while both low-class groups, CbWf e CbWm, produce the same phone with 50% and 43% RF, respectively;
9. 74% RF of the cartographical occurrences of the combination [ɹ]-[Ø] come from low class; and
10. when articulated by high-class feminine individuals, the combination [ɹ]-[ɹ]-[Ø] visible on the corresponding chart (s. Figure 2) occurs in the diatopic dimension in a more consistent way in the upper part—having the Tietê river as the spatial reference—of the Médio Tietê region as framed in the cartographic representation.

Though conclusions 6 and 7 reveal almost the same values concerning the retroflex approximant, the informant groups involved are diametrically opposite. The one group consists of high-class women. The other group consists of low-class men. As for conclusion 8, it is seen that women in general, more than men, prefer the variants [ɹ] e [Ø].

Fourthly and lastly, a correlational perspective between E1 and both diasexual and diagenational variables shows that:

11. both feminine and masculine young speakers are the ones to make the retroflex approximant become a relative norm (WfGI: 60% de FR; WmGI: 58%);

12. on account of old women, the phonetic zero almost turns into a relative norm with its 48% RF, a result that indicates a change in progress;
13. men are responsible for most of 67% RF of the combination [ɹ]-[Ø] present on the corresponding chart (s. Figure 3) throughout the network of localities except for Araçariçuama (P3) and Piracicaba (P10); and
14. women are responsible for most of 73% RF of the combination [ɹ]-[r]-[Ø] (20) present on the corresponding chart (s. Figure 3) throughout the network of localities except for Santana de Parnaíba (P1) and Pirapora do Bom Jesus (P2).

If the percentage values mentioned in conclusions 1 and 11—which concern a shared age group—are compared, and if the corresponding regular diatopic distribution of [ɹ] is taken into account, the consequence is that apparently neither the diastratic, nor the diasexual nor the diatopic variable exerts influence on determining the prevalence of the variant under consideration through its percentage values. Actually, the extralinguistic variable that seems to influence a greater usage of [ɹ] is diagenetic.

FIGUEIREDO JUNIOR, S. R. Róticos em coda externa de verbos no infinitivo no português caipira paulista. *Alfa*, São Paulo, v.66, 2022.

- *RESUMO: Com aporte teórico-metodológico da Dialectologia Pluridimensional (RADTKE; THUN, 1996; THUN, 2000, 2005 etc.), este artigo documenta e analisa a variação em coda externa de desinência infinitiva seguida por pausa no berço da cultura caipira paulista, na região chamada Médio Tietê. Com aplicação do Questionário Fonético-Fonológico do Atlas Linguístico do Brasil (COMITÊ NACIONAL DO PROJETO ALiB, 2001), os dados foram coletados in loco por Figueiredo Jr. (2019) em dez localidades: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariçuama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari e Piracicaba. Os inquiridos, totalizando 80, eram equitativamente mulheres e homens, jovens (18–36) e velhos (≥ 55), com baixa escolaridade (desde analfabetos até secundaristas incompletos) e alta (a partir de estudos superiores incompletos). As variantes documentadas são: aproximante retroflexa ([ɹ]), tepe alveolar ([r]), apagamento fonético ([Ø]) e vibrante alveolar ([r]). A análise baseia-se em estatística descritiva e cartas dialetológicas pluridimensionais, de que um conjunto relevante de conclusões decorre, entre as quais, de uma perspectiva ampla, a variável linguística estudada é saturada sobretudo pela aproximante retroflexa e, de uma perspectiva correlacional, o fator externo de influência no estabelecimento da prevalência da aproximante retroflexa não é, segundo evidências, nem diastrática, nem diagenética e nem diatópica, mas sim diagenética.*
- *PALAVRAS-CHAVE: dialectologia pluridimensional; geolinguística; fonética; português paulista; dialeto caipira; Médio Tietê; arquifonema /R/ em coda silábica.*

REFERENCES

- AMARAL, A. **O dialeto caipira**. São Paulo: Anhembi, 1920.
- BARRY, W. J. Another R-tickle. **Journal of the International Phonetic Association**, Cambridge, v.27, ns.1–2, p.35–45, 1997.
- BRANDÃO, S. F.; PESSANHA, D. B. S.; PONTES, S. P.; CORRÊA, M. O. Róticos na variedade urbana do português de São Tomé. **PAPIA**, São Paulo, v.27, n.2, p.293–315, 2017.
- CALLOU, D.; SERRA, C. Variação do rótico e estrutura prosódica. **Revista do GELNE**, Natal, v.14, n.esp., p.41–57, 2012.
- CÂNDIDO, A. **Os parceiros do Rio Bonito**: estudo sobre o caipira paulista e a transformação dos seus meios de vida. São Paulo: Duas Cidades, 2001.
- CHABOT, A. What’s wrong with being a rhotic? **Glossa: a journal of general linguistics**, Londres, v.4, n.1, p.1–24, 2019.
- COHEN, E.-G.; LAKS, L.; SAVU, C. The phonetics of Modern Hebrew rhotics. **Brill’s Journal of Afroasiatic Languages and Linguistics**, Leiden, v.11, n.1, p.28–48, 2019.
- COMIOTTO, A. F.; MARGOTTI, F. W. Uso dos róticos do português em contato com os dialetos italianos. **Acta Scientiarum. Language and Culture**, Maringa, v.41, n.2, p.1–9, 2019.
- COMITÊ NACIONAL DO PROJETO ALiB. **Atlas Linguístico do Brasil**: questionários 2001. Londrina: Ed. da UEL, 2001.
- FIGUEIREDO JUNIOR, S. R. **Atlas linguístico pluridimensional do português paulista**: níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê. Advisors: Manoel M. Santiago-Almeida, Harald Thun. 2019. 2.120 p. Thesis (Doctor in Lettres) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019.
- IBGE. **Cidades e Estados**: São Paulo. Brasília: IBGE, 2021. Available in: <https://www.ibge.gov.br/cidades-e-estados/sp.html>. Access on: 07 Feb. 2022.
- LADEFOGED, P.; MADDIESON, I. **The sounds of the world’s languages**. Oxford: Blackwell, 1996.
- LINDAU, M. The story of /r/. In: FROMKIN, V. (org.). **Phonetic linguistics**: essays in honour of Peter Ladefoged. Orlando: Academic Press, 1985. p.157–168.
- PATIN, C. /r/ in Washili Shingazidja. In: SPREAFICO, L.; VIETTI, A. (org.). **Rhotics**: new data and perspectives. Bolzano: Bolzano University Press, 2013. p.173–190.

RADTKE, E.; THUN, H. Neue Wege der romanischen Geolinguistik: eine Bilanz. *In*: RADTKE, E.; THUN, H. (org.). **Akten des Symposiums zur empirischen Dialektologie**. Kiel: Westensee-Verl., 1996. p.1–24.

ROMANO, A. A preliminary contribution to the study of phonetic variation of /r/ in Italian and Italo-Romance. *In*: SPREAFICO, L.; VIETTI, A. (org.). **Rhotics: new data and perspectives**. Bolzano: Bolzano University Press, 2013. p.209–225.

SCHILLER, N. O. The phonetic variation of German /r/. *In*: BUTT, M.; FUHRHOP, N. (org.). **Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen**. Hildesheim: Olms, 1998. p.261–287.

SCOBIE, J. M. (R) as a variable. *In*: BROWN, K. (org.). **Encyclopedia of language & linguistics**. Oxford: Elsevier, 2006. p.337–344.

SEBREGTS, K. **The sociophonetics and phonology of Dutch r**. Utrecht: LOT, 2015.

THUN, H. A dialetologia pluridimensional no Rio da Prata. *In*: ZILLES, A. M. S. (org.). **Estudos de variação linguística no Brasil e no Cone Sul**. Porto Alegre: Ed. da UFRGS, 2005. p.63–92.

THUN, H. (dir.). **Atlas lingüístico diatópico y diastrático del Uruguay**. Kiel: Westensee-Verl., 2000.

WIESE, R. The representation of rhotics. *In*: OOSTENDORP, M.; EWEN, E.; HUME, C. J.; RICE, K. (org.). **The Blackwell companion to phonology**. Oxford: Blackwell, 2011. p.711–729.

Received on July 19, 2020

Approved on June 14, 2021