



## FOUNDER EFFECT IN TUPIAN LANGUAGES<sup>1</sup> EFEITO FUNDADOR NAS LÍNGUAS TUPI

*Cilene Rodrigues*<sup>2</sup>

### ABSTRACT

Since Mayr (1954, 1963), it has been observed that territorial expansions may cause *founder effects*, reducing the genetic variability of the founder population. A similar effect has been reported within linguistic typology, as the phonemic inventory size of a language is reduced due to territorial dispersal of its speakers. Atkinson (2011) analyzes global present-day phoneme inventory size as a reflection of a serial founder effect caused by human exodus out of Africa during the Paleolithic period: the further a language is from eastern-southern Africa, the smaller its phonemic inventory is. Recent studies have shown that this founder effect may interact with other factors such as population size, language contact and isolation. In the present study, we analyze the phonemic effects of the dispersal of Tupi-Guarani people. Taking the basins of Madeira and Guaporé rivers to be the birthplace of Proto-Tupi (RODRIGUES, 1964), we investigated whether physical distance from Madeira-Guaporé is related with phonemic clines within Tupi-Guarani languages located in the southwest region of South America. We also analyzed whether population size is a significant factor, but it did not interact with phonemic inventory size. A variance test, however, showed that languages in the Madeira-Guaporé region present, among themselves, more vocalic variation than southwest Tupi-Guarani languages. This result is compatible with a founder effect. Contrasting with Tupi languages spoken in Madeira-Guaporé, southwest Tupi-Guarani languages are quite harmonious with respect to vowels, presenting two symmetrical series of nasal and oral vowels. This result and the general discursion we present here converse with studies on reconstruction of Proto Tupian languages, and with investigations on the mental representation of vowels and consonants (NESPOR et al., 2003). If vowels are markers of structural relations, we predict that there is more structural cohesion (less parametric variation) among Southwest Tupi-Guarani than among Tupi languages at Madeira-Guaporé.

**KEYWORDS:** Founder effect; Territorial dispersal; Genetic variation; Phonemic inventory; Vowels.

---

1 This paper is in honor of Aryon Dall'Igna Rodrigues, who devoted his life to understand native languages. As the reader will see, the research we are advancing here would not be possible without Rodrigues's contribution to linguistics. I was his student at UnB and, as I truly admired his work, I promised myself that I would do research on Brazilian native languages once I had grown older...I guess it is, thus, about time for me to start fulfilling my promise. I am also deeply thankful to Henrique Rajão/PUC-Rio, who is my *Virgilio* in the divine realms of biology and evolution, and to Marci Fillet Martins/Museu Nacional/UFRJ, my *Breatrice*, my consult on a myriad of issues related to paradise - Brazilian native language and culture. Also, I am in debt with Thomas Krahe/PUC-Rio and Monica Chaves/PUC-Rio, who used their mental powers and generosity to help me with statistics. A thank you note goes also to everybody who carefully read and comment on the final version of the paper, especially the reviewers of *Diadorim*. The best of doing research is the possibility of learning with and from others. Of course, all misunderstandings and misanalyses are my fault.

2 Pontifícia Universidade Católica - Rio de Janeiro



## RESUMO

Desde Mayr (1954, 1963), tem-se observado que as expansões territoriais causam *efeito fundador*: o grupo que se desgarra do original (i. e. grupo fundador) apresenta menor variabilidade genética. Efeito similar tem sido relatado na linguística: o número de fonemas de uma determinada língua se reduz como resultado de expansões territoriais. Atkinson (2011) analisa o tamanho do inventário fonêmico de diversas línguas modernas como reflexo da diáspora africana durante o período paleolítico: quanto mais longe uma língua estiver do leste/sudeste da África, menor será o seu inventário fonêmico (efeito fundador seriado). Estudos recentes sugerem que esse efeito pode interagir com outros fatores como tamanho da população e contato e isolamento linguístico. O presente estudo analisa os efeitos fonêmicos da dispersão territorial dos povos Tupi-Guarani. Tomando as bacias dos Rio Madeira e Guaporé na Amazônia como local de origem do Proto-Tupi (RODRIGUES, 1964), investigamos se há interação entre distância física deste ponto-origem e tamanho do inventário fonêmico das línguas Tupi-Guarani modernas faladas no sudoeste da América do Sul. Analisamos também se o número de falantes é fator significativo, mas não houve interação com o tamanho do inventário fonêmico das línguas consideradas. Teste de variância, no entanto, constatou que as línguas Tupi da região Madeira-Guaporé apresentam, entre elas, maior variação vocálica do que as línguas Tupi-Guarani do sudoeste. Esse resultado é compatível com o efeito fundador relatado acima. Em contraste com a línguas do Madeira-Guaporé, as línguas Tupi-Guarani do sudoeste são muito harmoniosas entre si, apresentando duas series simétricas de vogais nasais e orais. Este resultado e os fatos gerais aqui apresentados dialogam com análises de reconstrução de Protolínguas dentro do tronco Tupi, mas também com investigações sobre representações mentais de consonantes e vogais (NESPOR et al., 2003). Considerando que as vogais são marcadores de relações estruturais, lançamos a previsão de há menos variação paramétrica entre as línguas Tupi-Guarani investigadas do que entre as línguas Tupi faladas no Madeira-Guaporé.

**PALAVRAS-CHAVE:** Efeito fundador; Variação genética; Dispersão territorial; Inventário fonêmico; Vogais.

## Introduction

The United Nations has declared 2019 the international year of indigenous languages. This initiative is based upon the fact that the majority of endangered languages are indigenous. According to the *Ethnologue – Languages of the World* (LEWIS, 2009)<sup>3</sup>, there are around 7.000 living languages, and 95% of them have less than one million speakers (average of 600.000 speakers per language). Thus, most of the living languages are spoken by native people and, unfortunately, the future prospect for these languages is rather sobering. It is estimated that half of them will be lost by 2100 (NETTLE & ROMAINE, 2000; CRYSTAL, 2000), with the rate of language death being around 9 languages per year, as 1 language dies every 40 days (SIMONS, 2019).

According to the 2009 version of the *Ethnologue*, 473 languages are nearly extinct, most of them being located in the following countries:

**Table 1:** countries with the biggest number of nearly extinct languages

	Country	Number of nearly extinct languages
1	Australia	97
2	USA	74
3	Brazil	37
4	Papua New Guinea	31
5	Canada	19

(Based on information from *Ethnologue*. [https://www.ethnologue.com/16/nearly\\_extinct/](https://www.ethnologue.com/16/nearly_extinct/))

The fact that Brazil is in third place in the rank above is a highly worrisome situation

that calls for immediate awareness and actions from Brazilian authorities and academics (e.g. educators, linguists, anthropologists and social health researchers, among others). It is important to develop academic, social, cultural and political strategies to give native people support and motivation to keep their languages alive. Documenting native languages in detail, especially those on the path of extinction, is equally important, as it is a *sine qua non* condition to guarantee future research on fundamental duets such as nature & nurture, culture & biology, diversity & universals.

In this paper, we illustrate the importance of language preservation and documentation considering the so-called *founder effect* (MAYR, 1954, 1963). Acknowledging a reported positive correlation between phonemic reduction and territorial expansions (ATKINSON, 2011), I will focus on the Tupian languages spoken within South America, evaluating, as a working hypothesis, if there is a cline in the phoneme inventories of these languages as a reflection of their distance from the basins of the rivers Guaporé and Madeira (state of Rondônia/Brazil), understood to be the original birthplace of Proto-Tupi, the common ancestor of all Tupian languages.

Research on founder effects and its consequent decrease on genetic and linguistic variability is an important piece of the evolution puzzle, shedding light on our origins and also on the driving forces behind evolution of diversity. By pushing forward investigations on possible phoneme reduction among the Tupian languages, we want to understand how territorial dispersal affects language diversity, and what it shows us about evolution of language. We also want to call attention to the potential contributions of Brazilian native people to this line of research. In addition, we want to join efforts with Brazilian and non-Brazilian linguistics to comprehend Tupian languages and to raise awareness about the importance of preserving and documenting them.

We truly hope that the investigation presented in the following pages makes the point that native languages are valuable pieces of information about whom we all are.

The paper is organized as follows: in section 2, a brief discussion of the founder effect within genetics is given. The goal of this section is to inform the reader about the research environment in which the term *founder effect* was first coined. In section 3, we discuss Atkinson's (2011) parallelism between genetic decay and phonemic reduction as a consequence of the first terrestrial expansion of *Homo sapiens* exodus out of Africa. Atkinson's conclusion is that, similarly to what is observed in genetics, reductions on phonemic variability within a given population X reflect the physical distance separating X from southern-eastern Africa, the possible original departing point of humans, and possibly the birthplace of language. As we will discuss, Atkinson's correlation was further investigated on large samples of populations and languages, reaching some interesting conclusions about other factors, such as population size, language contact and isolation. In section 4, we bring into this discussion the territorial

expansion of the Tupian people. As presented in section 4.1, a genetic cline related to this expansion has already been reported. In section 4.2, taking into account the hypothesis that Proto-Tupi emerged on the Madeira-Guaporé basins, we present a sample of the research we are currently conducting, contrasting the size of phoneme inventories of Tupian languages spoken in the region Madeira-Guaporé (languages from Tupian families) with the size of phoneme inventories of the Tupi-Guarani languages spoken in the south of South America. The main goal of the present paper is to verify whether the dispersal of Tupi-Guarani from Amazon (Rondônia) to southwest caused a linguistic founder effect. Section 5 is dedicated to conclusions.

### **Founder effect in genetics**

Similarly, to the relation phoneme-allophones, genes, the constant units of heredity, have a number of alternate forms called alleles, which are responsible for most differences among individuals.

Humans (an example of diploid organisms) are composed by pairs of the same gene, one inherited from the father, the other from the mother. If these two genes have the same allele, the organism is homozygous. If, in contrast, the two genes have different alleles, the organism is heterozygous. In an ideal, stable population of diploid organisms, a quasi-equilibrium is attained (variance close to the mean 0.5), with heterozygosity guaranteeing a normal rate of allele variability within the group (see RIDLEY, 2004; FUTUYMA, 2005; MAYR, 2013). It is important to observe that allele variability is desirable as it increases the number of successful responses that a species can give to processes driven by natural selection and random drifts.

Natural selection is a blind evolutionary force that shapes the genotype of a species, and it plays only with the genetic material it has at hands. Thus, suppose that a gene *A* (color) has two alleles *A* & *a* (black & white), but only *A* (black) fits well with certain changes in the environment (pollution, for example), then a population that has a high frequency of *A* will have a better chance of surviving in the new environment<sup>4</sup>. Conversely, if allele *A* is not present in the population, the chances of survival decrease. In general, then, heterozygous populations will always have a better chance of adaption<sup>5</sup>.

Natural selection, however, is not the only force in the play of evolution. Events such as epidemics, natural disasters, lethal contacts with other groups or species, and/or migrations, can be equally powerful in shaping a species' genotype as they cause a random sampling of alleles. That is, as a consequence of one of these events, a population may go through a bottleneck<sup>6</sup>,

---

4 This is well illustrated by the change in color (white vs. melanic) in peppered moths (*Bristol betularia*), see Cook *et al.* (2012).

5 A great example of heterozygous advantage is the interplay between thalassemia and malarial resistance, see Cavalli-Sforza (2001).

6 A bottleneck happens when the size of a population remains very small for at least one generation.

experiencing a decrease in size, an actual reduction in the number of individuals. As a result, a drift occurs, randomly shifting the frequency (proportion) of alleles in that population.

Important to our topic of discussion, the effects of genetic drifts are stronger in small populations than in large ones. To see this, imagine that you have two sets of coins, one with 100 units and the other with only 10 units. If you flip the coins, it is likely that the ratio head-tail will deviate more from the point of equilibrium (50-50 ratio) in the 10-coin set than in the 100-coin set. The same rationale applies to genetic drifts: a stochastic allelic assortment to form a small new population is likely to under-represent the allele frequency of the original population. Also important is the fact that drifts have an accumulative effect over time: the changes in frequency of alleles (i.e. deviance from a point of equilibrium) observed in a given population will increase in magnitude in the next generations. Thus, drift causes an accumulated loss of genetic variability within a given population<sup>7</sup>.

One particular case of genetic drift is the so-called *founder effect* (MAYR, 1954, 1963) defined as “the establishment of a new population by a few original founders (in an extreme case, by a single fertilized female), which carry only a small fraction of the total genetic variation of the parental population” (MAYR, 1963). That is, founder effects are observed in cases of territorial expansions when interbreeding between the new and the original group is not possible. Taking into consideration dispersal of small groups, we can reason that these migrations can lead to bottlenecks, and, consequently, to genetic drifts, putting, thus, genetic equilibrium at risk. Founder populations usually display a reduction in heterozygosity, which reflects a cline in allelic richness, although reduction in allelic richness can be more expressive than reduction in heterozygosity given that allelic reduction can affect rare alleles, which does not really contribute to heterozygosity (BARTON & CHARLESWORTH, 1984; DLUGOSCH & PARKER, 2008; PIERCE et al., 2014; MARQUES & RENESTO, 2017).

*Hemiodus orthonops*, a small fish endemic from the Paraná-Paraguay basin, has recently migrated to the upper part of the Paraná River, founding new populations. Marques & Renesto (2017) analyzed the genotype of two founder populations, concluding that they both display a low mean of alleles per locus and an excess of homozygotes, which the authors take to be a reflex of founder events during the migration process.

Founder effects have also been intensively studied in cases of accidental or forced isolation of human populations, and, in many of these studies, it has been argued that the arbitrary allele selection observed in these events can foster the permanence of alleles that would be selected against in large populations. For example, the high rate of *Porphyria variegata* (severe reaction to barbiturate anesthetics) in Afrikaners (RIDLEY, 2004) is but one example. The modern populations of Afrikaners are mainly descended from a small number of immigrants

---

<sup>7</sup> Notice, however, that, given that drifts cause populations to differ genetically from each other, we can say that drifts cause an increasing in genetic variability between populations. Thus, drifts are responsible for diversity.

(Dutch, Germans and French) that arrived in South Africa in between 1652-1806. Among these immigrants, there was a Dutch couple that migrated from Holland in 1685-1688, carrying the rare *Porphyria-variegata* gene. Nowadays, around 30,000 Afrikaners have this gene, a much higher rate than that observed in Holland, and most of them can be genetically traced back to that couple. Examples can also be found in cases of isolation for political, social and/or religious reason. The Amish and Mennonite communities of North America are instances of social segregations, which led to founder effects. These groups were created in Europe in the 16<sup>th</sup> century, during the protestant reformation, for religious reasons. However, they were persecuted and some migrated to North America, where they kept themselves isolated, preventing genetic exchanges with outsiders via observation of restricted endogamy mating rules (PAYNE et al., 2011). Among the Amish community of Lancaster, state of Pennsylvania (USA), there is a large incidence of *Ellis-Van Creveld* Syndrome, a rare genetic disorder of skeletal dysplasia, which is autosomal recessive, meaning that both parents have to be carriers of the gene for a child to be affected by the syndrome. As people in these communities keep very good genealogical and health records, it is possible to recover the genetic relations among them, and a series of studies have focused on understanding why they present high rates of this otherwise rare syndrome. McKusick (2000) reports the existence of 50 *Ellis-Van Creveld* children, whose parents are both descendants from Samuel King and his wife, who migrated to Pennsylvania in 1744.

The birthplace of anatomically modern humans is motif of much debate (CAVALLI-SFORSA et al., 1993; CAVALLI-SFORSA 2001; RIDLEY, 2004), but there seems to be a consensus that, around 40.000 B.P.<sup>8</sup>. (Lower Paleolithic period), Europe was populated by waves of humans coming from eastern-southern Africa (CAVALLI-SFORSA et al., 1993; CAVALLI-SFORSA, 2001; RITO et al., 2019)<sup>9</sup>. This was probably the most impressive and the hardest territorial expansion that our species ever did. Having started a radial dispersal from eastern-southern Africa, around 100.000-70.000 B.P. (Upper Paleolithic period), little by little, small groups of humans reached Europe, Asia, Oceania and America. Data from paleoanthropology and archaeology are consistent with the monogenesis hypothesis that takes eastern-southern Africa to be the cradle of modern humans (RITO et al., 2019)<sup>10</sup>. Genetic evidence also favors this hypothesis. Africa (specially the eastern-southern area) is the place where genetic diversity is mostly accentuated in the world, even though the generic difference between the major geographically defined modern groups is very slight, 0.7. 93% of genetic variation is present in all human groups, the remaining 7% is responsible for all the differences

---

8 B.P. = Before present time.

9 There is no strong consensus about the specific region in Africa, as different evidence points towards different areas. However, most researchers accept that it was either eastern or southern Africa, but we still don't know which of the two. One possibility is that humans started in a region within the eastern area and moved in a radial fashion to south, north and west (see CAVALLI-SFORZA, 2001 and RITO et al., 2019).

10 Accepting a weaker interpretation, we might say that this evidence is at least consistent with polygenesis hypotheses, according to which eastern/southern Africa is the main region of human origin, but not the only one.

among us all (RIDLEY, 2004)<sup>11</sup>. These two facts (concentration of genetic diversity in Africa and the slightness of genetic differences among us all) advocate in favor of the hypothesis that all modern humans share a common African ancestor. It also tells us that the out-of-Africa exodus resulted in loss of genetic variation. Cavalli-Sforza (2001), considering the genetic distance (genetic differences) between African native people and native people from the other four continents (Europe, Asia, Oceania, and America), gives us the percentages on table 2. Based on these percentages, we can more or less infer the path we went through when we expanded from Africa: first we settled in Europe, then in Asia and, latter on, we moved to America and Oceania.

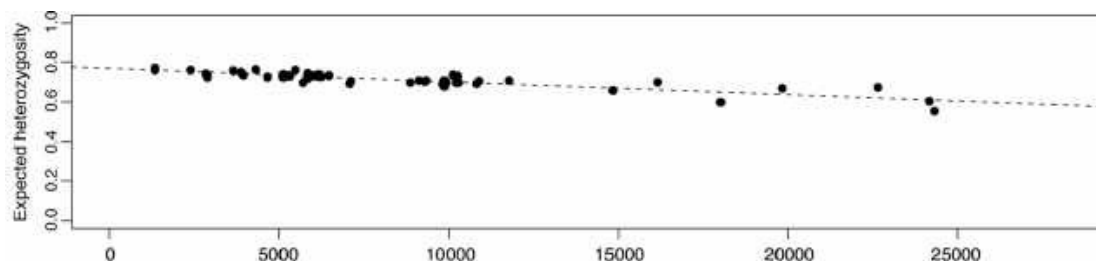
**Table 2:** genetic distance between Africa and the other continents

<b>Africa</b>	16.6	Europe
	20.6	Asia
	22.6	America
	24.7	Oceania

(based on Cavalli-Sforza, 2001: 52)

Several founder effects accompanied this admirable several-steps territorial dispersal. For example, heterozygosity decreases in a serial fashion as geographic distance from Africa increases, see figure 1 (RAMANCHANDRAN et al., 2005). That is, heterozygosity in a group X reduces in function of X geographical distance from Africa. Hence, American native people have less heterozygosity than Europeans, which in turn have less than Africans.

**Figure 1:** decay of heterozygosity plotted against geographic distance from Addis Ababa, Ethiopia



(Based on RAMANCHANDRAN et al., 2005)

Another example is the rate of type O blood. Among Native Americans, this rate is 98% (almost 100%), even though they decent from Asians that have only 50% of O. Thus, geneticists entertain the hypothesis that perhaps the first humans to reach America, crossing over the Bering Strait, were carriers of the O allele (CAVALLI-SFORZA, 2001).

As we will see in the next section, the human expansion from Africa might have also been accompanied by a serial founder effect within languages, promoting a cline in phoneme inventories.

<sup>11</sup> When we compare ourselves to other species, this low percentage becomes quite emphatic. See Ridley, 2004:365 for a comparison.

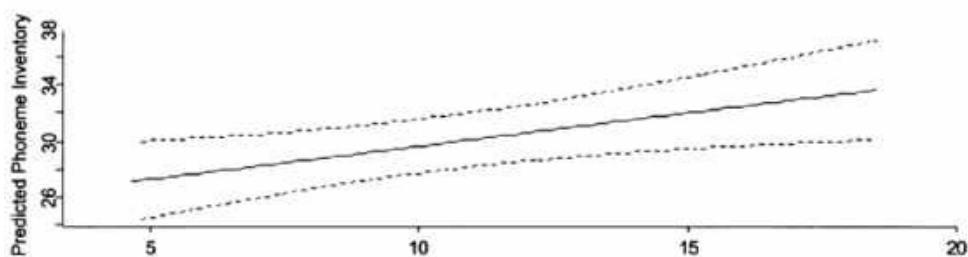
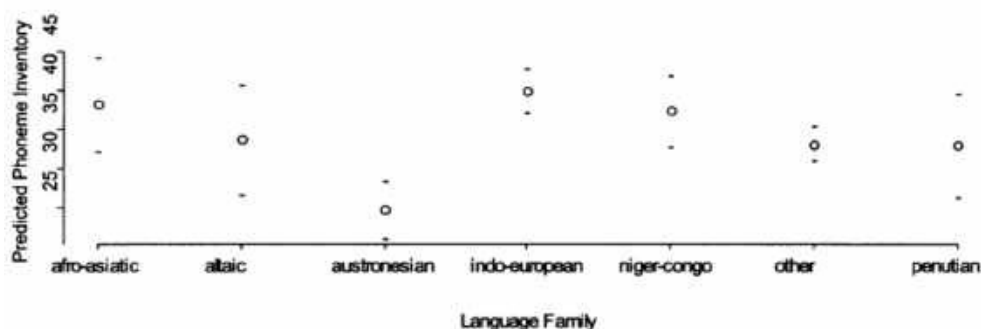
## Founder effect in linguistics

In language typology, a positive correlation between phoneme inventory size and population size has been observed. Languages spoken by large populations have more phonemes than languages spoken by small populations (TRUDGILL, 2005; HAY & BAUER, 2007).

Hay & Bauer (2007) verified this correlation statistically in a sample of 216 languages, including languages with a large number of speakers, as English (Indo-European) and Hindi (Indo-European), and languages spoken by “few” speakers, such as Basque (isolated), Diyari (Pama-Nyungan, Australian Aboriginal) Hixkaryana (Cariban, Amazon/Brazil)<sup>12</sup>. They considered various subgroups of phonemes. Vowel segments were divided in monophthongs and diphthongs, and monophthongs were further divided in basic monophthongs, which differ in quality only, and extra monophthongs, which differ in nonquality features, such as length and nasalization. Consonants were separated in plosives, fricatives, obstruents and sonorants.

Their results show a positive population-size-phoneme-inventory-size correlation: for any given language, the bigger the number of speakers, the bigger the phonemic inventory is. Thus, language families with a big number of speakers have more phonemes than language families with a small number of speakers, as shown in figure 2.

**Figure 2:** Increase in phoneme inventory size plotted against language family (first graphic) and against population size (second graphic)



(HAY & BAUER, 2007: 393)

Interestingly, as the *p* values in *table 3* show, when all phonemes are grouped together, a very strong positive correlation is obtained. However, as acknowledged by the authors, sonorant

<sup>12</sup> !Xu (!Xu, Africa) and Acooli (Nilo-Saharan, Uganda/Africa) were both eliminated from the sample as they show values more than four standard deviations above the mean. !Xu for total consonants and Acooli for total monophthongs



consonants, as well as the group of all vowels/monophthongs are less affected by fluctuations in population size.

**Table 3:** Spearman correlation between means of language family, population and phoneme inventory size

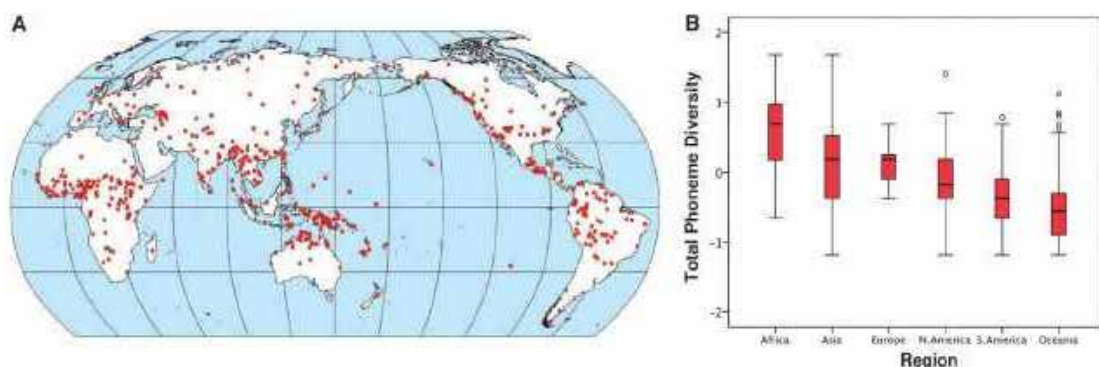
CORRELATION WITH	RHO	$p <$
All phonemes	.46	0.003
Basic monophthongs	.47	0.002
All monophthongs	.2	0.2
Diphthongs	.53	0.001
Plosives	.33	0.05
Fricatives	.53	0.001
Sonorants	.24	0.13
Obstruents	.43	0.005
All consonants	.45	0.005

(HAY & BAUER, 2007: 394)

Hay & Bauer (2007) did not really offer any explanation for this reported correlation, and Trudgill (2005) reasoned the size of a phoneme inventory might actually be related to the number of neighboring a language has, with languages in contact displaying more phonemes. However, it has been recently argued that languages with many neighbors present less phonemic variability than isolated ones (CREANZA et al., 2015).

Atkinson (2011) analyzes global present-day phoneme inventory (consonants, vowels and tones) size as a reflex of a serial founder effect caused by the human expansion from Africa during the Paleolithic period (see section 2). A statistical analysis was conducted on a sample of 504 modern languages from different families, located in different continents (figure 3/A). The data were extracted from *WALS – The world Atlas of Languages Structures* (DRYER & HASPELMATH, 2005). The results are in accordance with the conclusion that population size is a prediction of phoneme inventory size. They also indicate a positive correlation between phoneme reduction and physical distance from Africa: the further a language is from eastern Africa, the smaller its phoneme inventory is. As shown in figure 3/B, the highest level of phonemic diversity is in Africa, whereas the lowest is in South America and Oceania. Thus, the last continents to be occupied by humans are the ones that exhibit less phonemic diversity.

**Figure 3:** A: location of the 504 sampled languages, B: overall phoneme diversity plotted against region



(ATKINSON, 2011: 333)

Atkinson (2011) argues that the obtained results are in accordance with a serial founder effect parallel to that related to decay of genetic diversity (see table 2). The author's line of reasoning is that the human exodus from Africa resulted in a reduction of population size, bottlenecks, that led to imperfect phonemic transmissions, and, consequently to serial phonemic clines.

Many researchers reviewed Atkinson results. Cysouw et al. (2012), for example, tried to replicate them, using UPSID database -*UCLA Phonological Segment Inventory Database* - (MADDIESON & PRECODA, 1990) together with tone inventories from WALS<sup>13</sup>. Contrastively, their statistical results point towards two origin places, eastern Africa and Caucasus. Also, the authors observe that the population-size-phonemic-inventory-size correlation is obtained for large populations only. It reaches strong significance (at the 5% level) only when languages with a hundred thousand speakers or more are included in the sample. Thus, it is very likely that the purported correlation would not hold for the Paleolithic hunter-gatherer groups that left Africa (see also CREANZA et al., 2015; FLEMING, 2017 on this matter). Cysouw et al. also wonder why phonemes, as opposed to other properties of language, are sensitive to population size (bottlenecks) and founder effects. The authors show that, if Atkinson comparative methodology is applied to other language properties, many dispersed sites are located as possible points of origin. For example, Uradhian languages (Australia) exhibit the largest number of reduplication processes, while Yupikian languages (Eurasian-North American border) display the most complex syllable structure<sup>14</sup>.

Taking phoneme inventory sizes as the basis for an analysis of language evolution, Perreault & Mathew (2012) conclude that languages spoken by a large number of individuals increase their phoneme inventories over time. The authors developed an ingenious model for calculating the rate at which languages accumulate phonemes. They took into consideration the first human occupation of the region spanning from the coast of southern India to the Malay Peninsula, which was called population A. Around 65.000-45.000 B.P., A dispersed, at the same time, over the southeast of Asia, founding population B, and over Andaman Islands/Indian Ocean, founding population C. Languages spoken by modern B & C differ in two aspects: B languages have a much larger number of speakers than C languages; and 41.2 is the average number of phonemes in B languages, while 24 is average in C languages. Assuming that the phonemic inventory size of C languages did not really change since the region was colonized, Perreault & Mathew took 24 to be the representative number of phonemes of the language that came from Africa with population A. Hence, based on (a) the elapsed time since colonization

---

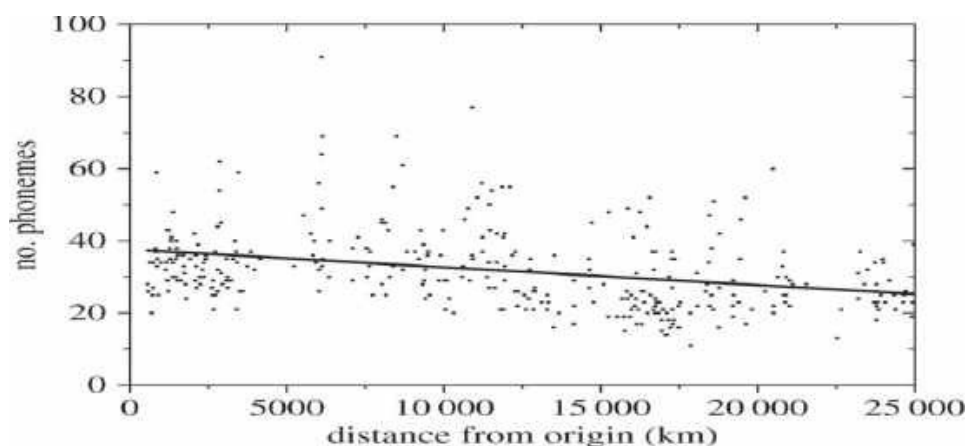
<sup>13</sup> According to the authors, WALS database puts too much weight on the numbers of vowels and tones, which might have pushed Atkinson's statistical results towards southwestern Africa as the original common place for modern languages.

<sup>14</sup> See Cysouw's et al. supplemental material.

of B & C (65.000-45.000 B.P.); (b) the average of the phoneme inventory of modern language spoken in B (41.2), and (c) the approximate phoneme inventory size for first language A (24), the authors calculated that the phoneme inventories of languages B increased at rate between 0.26 and 0.38 per 1.000-year period.

Considering all the facts above, and based on statistical analysis of 366 languages whose phonemic inventories were recorded in both WALS and UPSID, Fort & Pérez-Losada (2016) revisited Atkinson's results and conclusions. Their simulations also resulted in a phonemic cline that is compatible with a serial founder effect (see figure 4). However, they observe that these results hold only under three assumptions posited by Perreault & Mathew (2012): (a) phonemes accumulate in a rate between 0.26 and 0.38 per 1.000-year period, (b) at the onset of the out-of-Africa dispersal, languages had small phonemic inventories (around 11 segments); (c) the rate of phoneme accumulation depends on population size.

**Figure 4:** number of phonemes (366 languages) plotted against distance (Km) from



Africa (FORT & PÉREZ-LOSADA, 2016: 3)

Another interesting observation on the endless discussion about the way language might have evolved is offered by Fleming (2017). Fleming answers Cysouw's et al. worries about Atkinson's emphasis on phonemes and on Eastern/Southern Africa as the origin point, by considering consonantal clicks. Consonantal clicks occur as phonemes only among Khoisan languages, which are found across eastern and southern Africa<sup>15</sup>. The largest phonemic inventories are found within these languages and this is entirely due to the presence of clicks. !Xóõ (Tuu family), for example, has 119 phonemes, of which 80 are clicks (MILLER, 2011). Following Hockett's (1960) observation that duality of patterning was one of the last features of language to emerge, Fleming reasons that a human protolanguage was probably a monoplanar system, which means it was a communication system in which minimal units of form are directly

<sup>15</sup> Greenberg (1963) grouped these languages together as Khoisan, but they are not related genealogically, their common feature being the presence of consonantal clicks. (GÜLDEMANN, 2007). Maybe Khoisan can be understood as an areal language family in the sense of Campbell (2015).

mapped into meaning. Given that a monoplanar system requires a large inventory of minimal units in order to be able to refer to a good amount of entities and situations, clicks might have been part of human protolanguage because they can be simultaneously produced with other segments, multiplying phonemic distinctions. Hence, the fact that these elements are still part of the phonemic inventories of Khoisan languages might be evidence that language evolve in Africa, and, in the latest stages of its evolution, when duality of patterning was added, the clicks from the previous stage were kept. As the small hunter-gatherer groups of humans distanced themselves from Eastern-Southern Africa, they did not have contact with other click languages and, consequently, did not keep clicks in their phonemic repertoires, which were consequently reduced.

At any rate, before we go on, a proviso is in order. Studies and findings on genetics and linguistics do not walk hand in hand, especially in evolution, where different types of forces are at play. Not always, there is a match between genetics and linguistics. For example, Ethiopians are genetically related to Africans, but they speak languages that are Afroasiatic, which occur in the north of Africa and in the Middle East, where people are genetically Caucasoids (CAVALLI-SFORZA, 2001). This is due to the fact that genetic information is passed only vertically, from parents to offsprings, while linguistic information is arguably passed vertically and horizontally between genetically unrelated individuals. Also, language changes are much faster and more common than genetic changes, mutations. That is, genetic properties are more stable than language properties, although not all properties of language are equal with respect to this. Grammatical features are presumably more stable than phonemic features. As for Atkinson's hypothesis, it should be observed, as Sproat (2011) and Fort & Pérez-Losada (2016) did, that the reported phonemic reduction can result from a serial founder effect either because small populations lose phonemic diversity or because big populations gain it.

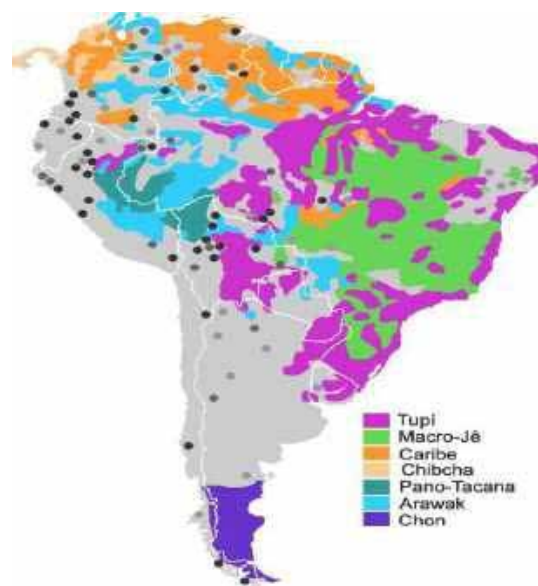
In what follows, we present part of our research on phonemic diversity within South American languages. Here we show that the dispersal of the Tupi-Guarani people from Amazon to the southern region of South America led to loss of variability in the vowel system. This is the first time this linguistic research is conducted and we believe it can add to the general discussion presented above, as reconstruction of the territorial dispersal of South America natives is relatively easier to be traced when compared to migrations during the Paleolithic period, like the human exodus out of Africa. Also, within the same language stock, all languages are genealogically related. This allows us to observe in a more direct and transparent way the effects of territorial dispersal over a phonemic inventory, steering away from big differences among language families/stocks. In addition, given that almost all South American native languages are spoken by small groups, we can put aside questions about correlations between population size and phoneme inventory size (SPROAT, 2011 and FORT & PÉREZ-LOSADA,

2016). Therefore, South American native languages might be good models for understanding the effects of territorial dispersal on human language.

### The territorial dispersal of the Tupian people within Brazil

Before the European colonization of South America, the Amerindians were already spread out, as shown in figure 5<sup>16</sup>. Notice that Tupi is one of the biggest language stocks, and it is largely concentrated within Brazil.

**Figure 5:** Native languages spoken in South America before the European colonization







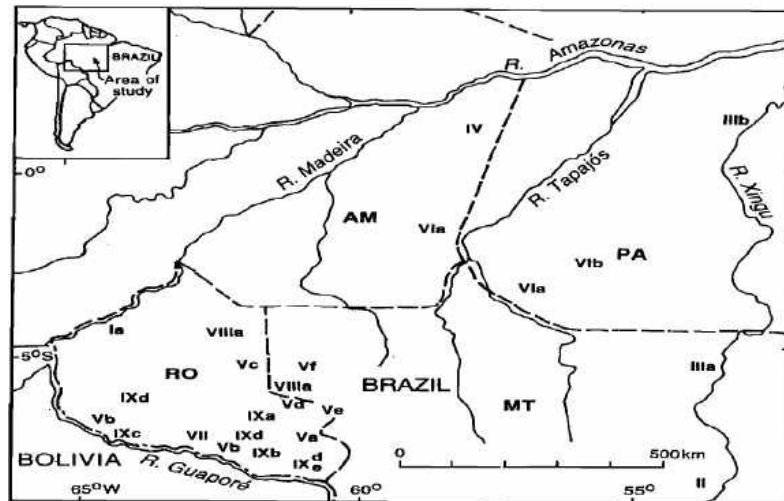
(from [https://pt.wikipedia.org/wiki/L%C3%ADnguas\\_do\\_Brasil](https://pt.wikipedia.org/wiki/L%C3%ADnguas_do_Brasil))

Tupi branches in 10 families: Arikén, Awetí, Juruna, Mauwé, Mondé, Mundurukú, Puruborá, Ramaráma, Tuparí and Tupi-Guarani (RODRIGUES, 1986, 2007a, RODRIGUES & DIETRICH, 1997). The map bellow (Figure 6) shows the approximated location of these families and their languages, except Tupi-Guarani, which we will discuss later.

Tupi branches in 10 families: Arikén, Awetí, Juruna, Mauwé, Mondé, Mundurukú, Puruborá, Ramaráma, Tuparí and Tupi-Guarani (RODRIGUES, 1986, 2007a, RODRIGUES & DIETRICH, 1997). The map bellow (Figure 6) shows the approximated location of these families and their languages, except Tupi-Guarani, which we will discuss later.

<sup>16</sup> For a recent overview of the history of Brazilian native languages, see Duarte (2016).

**Figure 6:** approximated locations of Tupian languages. I **Arikém** (a. Karitiana, b:  Arikém), II **Awetí** (Awetí), III **Juruna** (a. Juruna, b. Xipaya, c.  Minitawá), IV **Mawé** (Mawé), V **Mondé** (a. Mondé, b. Aruá, c. Gavião, d. Suruí, e. Cinta-larga; f. Zoró), VI **Mundurukú** (a. Mundurukú, b. Kuruáya), VII **Puruborá** (Puruborá), VIII **Ramaráma** (a. Karo, b.  Urumi), IX **Tuparí** (a. Tuparí, b. Wayoró, c. Menkéns, d. Makuráp, e. Sakirabiát,  Kepkiriwát)



(RODRIGUES, 2007, p. 108)

Five of these families are located either within Rondônia or in an area nearby (GABAS, 2006, RODRIGUES 2007a). Some languages of the Kawahib complex, a cluster belonging to Tupi-Guarani, are located in this region as well. Based on this, Rodrigues (1964) proposed that the Proto-Tupi language emerged somewhere at the basins of Madeira and Guaporé rivers (Henceforth MGR- Madeira-Guaporé Region).

In general, there is a consensus from history, archeology, anthropology, linguistics and genetics with respect to the broad region where Tupi first emerged. The bulk of evidence suggests that the Tupian motherland was located in the central-western Amazon, in the area bounded by the rivers Amazon on north, Tocantins on east, Madeira and Guaporé on west, Guaporé on south, (See NOELLI (1996) for a detailed discussion). Archeological approaches (LATHRAP, 1970; BROCHADO, 1984) consider that the original point was in Central-Amazon, on the confluence of Amazon and Madeira Rivers. Lathrap also hypothesized that Tupians, pressed by the Arawak people, migrated towards south. Linguists, on the other hand, concluded that Proto-Tupi originated at MGR (RODRIGUES, 1964). Notice that these two approaches (Archeology (Lathrap/Brochado) and Linguistics (Rodrigues)) are not totally incompatible. If Lathrap's hypothesis is right, a migration southwards would place the Tupian people around MGR.

Also, based on glottochronology, it is estimated that Proto-Tupi emerged 5.000 B.P. (RODRIGUES, 1964; URBAN, 1992, 1996), and the first Tupi languages (which are language families nowadays), including Proto-Tupi-Guarani, started branching out around 3.000 B.P.

The Tupi-Guarani family is special in many ways. First, it is the biggest branch of the Tupian tree, being composed by approximately 40 languages, which were divided by Rodrigues (1985) in eight different groups based on grammatical similarities (See also RODRIGUES & CABRAL, 2002; RODRIGUES & CABRAL, 2012; GABAS, 2006; MICHAEL et al, 2015, among others). Table 4 presents the Tupi-Guarani languages organized in the eight groups proposed by Rodrigues (1985).

**Table 4:** Tupi-Guarani languages

Família Tupi-Guarani							
Ramo I	Ramo II	Ramo III	Ramo IV	Ramo V	Ramo VI	Ramo VII	Ramo VIII
Guaraní Antigo	Guarayu	Tupinambá	Assuriní	Kayabí	Parintintín	Kamaiurá	Takunyapé
Guaraní Mbya	Sirionó	Língua Geral Paulista	Tapirapé	Assuriní Xingu	Tupi-Kawahíb		Emerrillon
Xetá	Jora (Bolívia)	Nheengatu (Língua Geral Amazônica)	Ava Canoeiro	Araweté	Apiaká		Ka'apor
Nandeva	Cocama		Suruí				Wayampí
Kaiwá	Cocamilla		Parakanã				Amanayé
Guarani Paraguaio	Omagua		Guajajára				Anambé
Guayakí			Tembé				Turiwara
Tapieté							Guajá
Chiriguano (Guarani da Bolívia)							
Izoceño							

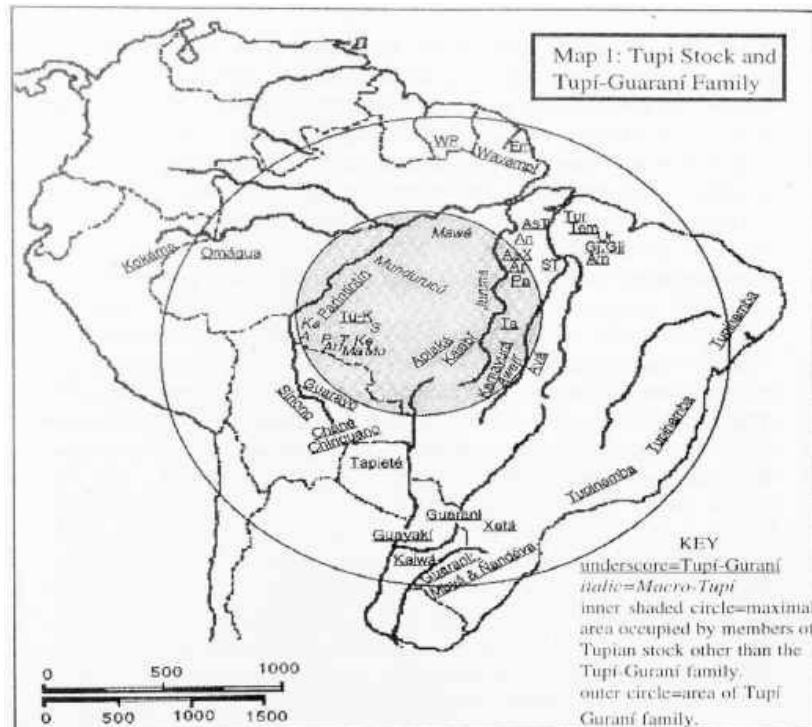
(DUARTE, 2016)

As table 4 shows, the Tupi-Guarani languages are spread in a vast area within South America. If the center of dispersal was MGR (around 3.000 B.P.), the Tupi-Guarani people spread out in a radial fashion towards north, south and east. Archeological data indicate that around 2.000 B.P. they had already settled down in the south of Brazil, in the states of Paraná and Rio Grande do Sul (NOELLI, 1996).

Urban (1996) presents an interesting discussion on the territorial expansion of the Tupi-Guarani people. The reasons for why they migrated are unknown. As Urban puts it, it could have been for many reasons: search for a new ecological niche, and introduction of new cultigen in the agriculture system, use of new transportation mode (canoe), new cosmology (search for an earthly paradise), new orientation in terms of space. Although we don't know the reasons

for their displacement, we know they were fast and their expansion took them far away from their birthplace. As the map below shows (figure 7), their total rate of expansion is about four times that of other Tupi groups that migrated as well (e.g. Juruna, which is located in the State of Mato Grosso).

Figure 7: Areas occupied by Tupi-Guarani and other Tupi groups



(URBAN, 1996: 66)

An interesting cluster of Tupi-Guarani languages is the so-called Kawahíb complex (or Tupi-kawahíwa – RODRIGUES & CABRAL, 2012; AGUILAR, 2015, MARÇOLI, 2018). Among these languages, Juma, Parintintin, Tenharim, Uru-eu-uau-uau, Amondawa, Karipuna, Diahói are still located within Amazon<sup>17</sup>. Nevertheless, according to Nimuendajú (1948) (see also ALMEIDA & NEVES, 2015), the Kawahíb were not always where they currently are, they migrated (back) to their current places during the colonial period. Hence, they fit the migratory behavior of Tupi-Guarani.

All in all, in what follows we will show that the territorial dispersal of the Tupi-Guarani people caused founder effects within genetics and linguistics.

17 Juma (4 speakers) Parintintin (150) and Diahói (90) are located on the basins of Maici and Madeira; Tenharims (130) are in the upper part of Marmelos River and on the basins of the Iguapé Preto; Uru-eu-uau-uau (62) and Amondawa (65) and Karipuna (14) are located in central Rondônia, although isolated from each other. (See SAMPAIO, 1998; RODRIGUES, 2013). A demographic data from Siasi/Sesai (2014) ([https://pib.socioambiental.org/en/Table\\_of\\_Indigenous\\_Peoples](https://pib.socioambiental.org/en/Table_of_Indigenous_Peoples)) indicates a larger population among the Karipunas (55).



## Tupian populations: genetic founder effect

Genetic data on Amerindians are scarce. However, there is enough evidence that local (within the continent) ancient migrations caused a genetic founder effect, with founder populations presenting less genetic variability.

Based on a statistical analysis of allelic-variability markers on 22 native Amerindians populations, Yang et al. (2010) reports a serial continent-wise decrease in variability from north to south. Populations located in the south present a gradual reduction in variability, when compared to those located in the north. Studies with Brazilian Amerindians have reached similar results. Ramallo et al. (2013) analyzed variation among different groups of Jê and Tupi-Guarani, concluding that Tupi-Guarani populations, as opposed to Jê groups, present an isolation-by-distance genetic pattern. More recently Santos et al. (2015), focusing on Tupi populations, concluded that the territorial dispersal of Tupians caused a founder effect<sup>18</sup>.

Considering different types of genetic markers inherited in patrilineal (Y-chromosome) and matrilineal (mtDNA) way, Santos et al. (2015) screened data from different Tupi populations looking for variability. Taking MGR to be the birthplace of Tupi, the authors then contrasted data from Tupian groups located in MGR (Zoró, Gavião, Suruí, Cinta Larga, Karitiana) with data from groups located outside MGR (Mundurukú, Asuriní do Trocará, Asuriní do Coatinemo, Araweté, Urubú-Ka'apor, Parakanã, Awá-Guajá, Wayampi, Zoé, Aché, Guarani). Thus, their analysis contemplated data from populations that had expanded to south (e.g. Guarani, located all over south of Brazil) and to north (e.g. Wayampi, located in the border between Brazil and French Guiana) and northeast of Brazil (e. g. Urubú-Ka'apor, located in the State of Maranhão). Their comparison shows that decays in allelic frequencies are one of the hallmarks of the Tupian expansion. Table 5 below shows heterozygosity estimations for MGR and for NonMGR populations<sup>19</sup>.

**Table 5:** Variation in genetic markers in MGR and NonMGR Tupian Population

Markers	MGR	No-MGR
Classics <sup>1</sup>	0.193	0.172
STR <sup>2</sup>	0.683	0.632
Y-STR <sup>3</sup>	0.253	0.164
Alu insertions <sup>4</sup>	0.235	0.213
mtDNA sequence <sup>5</sup>	0.865	0.845

(SANTOS *et al.*, 2015:4)

Also, considering variation in the sequence of key positions within the nucleotide of mitochondria's DNA, haplogroups A, B, C, D, X, Santos' et al. results indicate that D is the most expressive haplogroup within Tupians. However, its highest frequency is found in MGR populations, over 60%. NonMGR Tupians presented frequencies below 30%, except for Mundurukú, whose frequencies is 55%. As the authors pointed out, Mundurukú is located near

<sup>18</sup> See also Bisso-Machado et al., (2012).

<sup>19</sup> The genetic markers included on table 4 are used to detect levels of heterozygosity in populations in general.

MGR. Hence, we may conclude that Tupi populations outside MGR display a haplogroup D reduction.

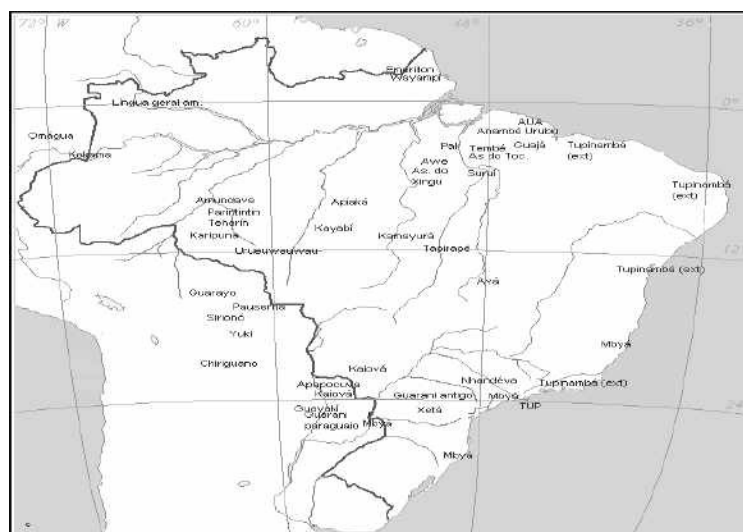
Importantly, Santos' et al. results point towards a recent reduction in both MGR and NonMGR populations, which the authors take to be a cumulative effect of different events such as reduction in natural resources and the European colonization process. However, their results also suggest that in ancient times Non-MGR groups suffered depopulations (bottlenecks), whereas MGR groups experience population growth.

### Tupian Languages: phonemic founder effect

Our research concentrates on verifying whether the ancient territorial expansions of Brazilian native populations led to a serial founder effect on language. In particular, we investigate whether these expansions induced a phonemic reduction similar to that detected by Atkinson related to the *Homo sapiens* out-of-Africa exodus. In this paper, we concentrate on the south territorial dispersion of Tupi-Guarani languages.

As already presented (section 4), it is assumed that the Proto Tupi-Guarani branched off from Tupi around 3.000 B.P. Around 2.000 B.P., Tupi-Guarani tribes were already settled down in the southernmost part of Brazil, in the states of Paraná and Rio Grande do Sul, but, as shown in the map (figure 8 below), they also settled down in Bolivia, Argentina and Paraguay. The distance between Rondônia, where MGR is located, and Rio Grande do Sul, the southernmost state of Brazil, is approximately 3.000 Km. Thus, the question is: did this long distance migration caused phonemic reductions that can still be captured in modern Tupi- Guarani languages?

**Figure 8:** location of Tupi-Guarani languages in South America



(MELLO, 2000)

Our working hypothesis is that it did. To test this hypothesis, we compare the phonemic inventories of Tupian languages located at MGR (TupiMGR) with the phonemic inventories of Tupi-Guarani languages located below Rondônia, in south of Brazil and in Bolivia, Argentina and Paraguay (SouthTG). These languages will be listed below. Notice that the migration routes that SouthTG people took is not completely understood. One hypothesis is that two migratory waves left MGR: the first went to Bolivia and the second one went further down towards the rivers Paraná and Uruguay (RODRIGUES, 2007a).

## Material and Methods

As already stated, we investigated the Tupi-Guarani territorial expansion towards southwest (figure 9)<sup>20</sup>, and we compared two groups of languages: TupiMGR and SouthTG.

**Figure 9:** territorial expansion of Tupi: southwest: Guarani; southeast: Tupinambá



(KNEIP & MELLO, 2013:21-23).

The separation between Tupi and Tupi-Guarani is entirely due to the comparison we chose to make. Importantly, the *kawahíb* complex, which is part of the Tupi-Guarani family, was removed from our sample. They are spoken at or nearby MGR, but they haven't been always there. As discussed in section 4, it is understood that the *kawahíb*s moved from and then back to MGR. Hence, they were removed from our sample, as their migratory history is unclear.

Apart from the exclusion of *Kawahíb* languages, other Tupi languages were selected to compose our sample based entirely on their locations and on the availability of information on their phonemic inventory, location and number of speakers. Data on phonemic inventories

<sup>20</sup> As Figure 9 shows, Tupi-Guarani went to Uruguay. However, no native language is spoken in Uruguay at present.

were extracted mainly from the database *SAPhon – South American Phonological Inventory Database*, (V1.1.4 <http://linguistics.berkeley.edu/~saphon/en/>). Whenever necessary, other sources were used as well. The *Ethnologue* database (<https://www.ethnologue.com/> - LEWIS, 2009) was used to verify information on location and number of speakers. Rodrigues (2013) was heavily used for this purpose as well.

The phonemic inventory of each sampled language was composed by consonants and vowels, which were analyzed separately. Also, vowels were divided in oral and nasals. Allophonic variations, tones and pitch accents were disregarded, as we could not find information for all the languages we sampled. Tones and pitch accent do not occur in Tupi- Guaraní languages. However, they have been shown to occur in other Tupi languages, such as Karitiana (STORTO & DEMOLIN, 2005) and Gavião of Rondônia (MOORE & MEYER, 2014). Also differences between dialects and languages were not taken into consideration because the distinction is not clear and our sample is small.

Our TupiMGR sample was composed by 10 languages: Karitiana, Gavião, Suruí/ Paitér, Káro, Makuráp, Akuntsú, Tuparí, Sakirabiá, Wayoró and Puruborá. Table 6 presents the data/ variables we compared in our analysis: place of location, number of phonemes (consonants and vowels) and number of speakers. It also indicates the source of information in each language.

**Table 6:** MGR languages: family, number of phonemes (consonants and vowels), number of speakers and source of information

<b>TUPIMGR</b>				
<b>Language</b>	<b>Phoneme inventory</b>		<b>Estimated number of Speakers</b>	<b>Source of information</b>
	<b>Consonants</b>	<b>Vowels</b>		
Karitiana (Arikém)	11	Oral: 10 Nasal: 10	320	- SAPhon - Ethnologue - Rodrigues (2013)
Gavião (Mondé)	21	Oral: 5 Nasal: 5	460	- SAPhon - Ethnologue - Rodrigues (2013)
Suruí Paitér (Mondé)	19	Oral: 5 Nasal: 5	1.000	- SAPhon - Ethnologue - Rodrigues (2013)
Puruborá (Poruborá)	14	Oral: 7 Nasal: 6	50	- Ethnologue - Rodrigues (2013) - Galucio et al. (2015)
Káro (Ramaráma)	14	Oral: 7 Nasal: 4	184	- SAPhon - Ethnologue - Rodrigues (2013)
Makuráp (Tuparí)	10	Oral: 5 Nasal: 5	380	- Ethnologue - Rodrigues (2013) - Braga (1992)

Akuntsú (Tuparí)	11	Oral: 6 Nasal: 5	6	- SAPHon - Ethnologue - Rodrigues (2013)
Tuparí (Tuparí)	14	Oral: 10 Nasal: 5	430	-SAPHon - Ethnologue - Rodrigues (2013)
Sakirabiá (Tuparí)	16	Oral: 5 Nasal: 5	85	- SAPHon - Ethnologue - Rodrigues (2013)
Wayoró (Tuparí)	14	Oral: 10 Nasal: 10	100	- SAPHon - Ethnologur - Rodrigues (2013)

The SouthTG sample was also composed by 10 languages: Kayowá (Mato Grosso do Sul/Brazil), Nhadeva (Mato Grosso do Sul/Brazil), Xetá (Paraná/Brazil), Mbyá (Rio Grande do Sul/Brazil), Chiriguano Izoceño (Bolivia), Chiriguano Chané (Bolivia), Sirionó (Bolivia), Paraguayan-Guarani (Paraguay), Tapieté (Argentina). These are languages from groups I and II of Rodrigues' (1985) classification and from subgroups Southern and Guaranian in Michael's et al. (2015) categorization<sup>21</sup>.

**Table 7:** SouthTG languages: family, number of phonemes (consonants and vowels), number of speakers and source of information<sup>22</sup>

SOUTH TG <sup>23</sup>				
Language	Phoneme inventory		Estimated number of Speakers	Sources of Information
	Consonants	Vowels		
Kayowá (Tupi-Guarani)	14	Oral: 6 Nasal: 6	15.000	- SAPHon- Ethnologue - Rodrigues (2013) - Dietrich (2010)
Nhadeva  (Tupi-Guarani)	17	Oral: 6  Nasal:6	570	- SAPHon - Ethnologue - Rodrigues (2013) - Costa (2007) - Mello (2000) - Dietrich (2010)
Xetá  (Tupi-Guarani)	13	Oral: 6  Nasal:5	3	- SAPHon - Ethnologue - Rodrigues (2013) - Mello (2000) - Vasconcelos (2008)
Guarani Mbyá	14	Oral: 6	7.000	- SAPHon

21 Rodrigues's classification is based on phonological features and Michael's et al. proposed grouping is based on lexicostatistics

22 Number of speakers is hard to measure. Here they are estimated for the whole population, not for groups/villages.

23 We will not discuss it here, but Ivo (2018) argues that all SouthTG languages spoken in Brazil present

(Tupi-Guarani)		Nasal6		- Ethnologue - Rodrigues (2013) - Mello (2000) - Maia et al. (2019)
Chiriguano Izoceño (Tupi-Guarani)	12	Oral: 6 Nasal: 6	51.230	- SAPHon - Ethnologue
Chiriguano Chané (Tupi-Guarani)	12	Oral: 6 Nasal: 6	2.440	- SAPHon - Ethnologue
Guarayu (Tupi-Guarani)	17	Oral:6 Nasal:6	5.933	- SAPHon - Ethnologue
Sirionó (Tupi-Guarani)	13	Oral: 6 Nasal: 6	650	- SAPHon - Ethnologue - Gasparani (2012)
Paraguayan Guarani (Tupi-Guarani)	18	Oral: 6 Nasal: 6	5.850.000	- SAPHon - Ethnologue
Tapieté (Tupi-Guarani)	15	Oral: 6 Nasal:6	750	- SAPHon - Ethnologue - González (2005)

Based on information available on *Ethnologue* on location of the Tupi-Guarani languages above, we used *Google maps* to measure the distance in kilometers between MGR and each of the SouthTG languages we sampled. We took the city of Cacoal in Rondonia as the origin point. Suruí/Paitér (table 6) is located in the municipality of Cacoal. Thus, we measured the physical distance between the Suruis and the SouthTG populations<sup>24</sup>. At *google maps*, we chose the “walking” parameter in order to simulate a path similar to that of the SouthTG people.

The comparisons consisted of verifying statistically if distance from MGR interacts with phonemic inventories size, where consonants and vowels were first considered separately and then together, forming a single group. In a similar fashion, we also investigated the existence of statistical interactions between population-size and phonemic-inventory-size.

---

the same consonantal system. If this is right, the number of consonants recorded on table 7 is not correct. In our research, to avoid biased choices of data sets resulting in either false positives or false negatives, we used, whenever possible, the same source of information - *SAPHon*.

24 Kayowá 2.153Km

Nhandeva 2.324Km

Xeta 2.153Km

Mbyá 3.058Km

Chiriguano Chané 1.239Km

Chiriguano Izoceno 1.239Km

Guarayu 1.263Km

Sirionó 1.474Km

Paraguayan Guarani 2.077Km

Tapiete 2.172Km

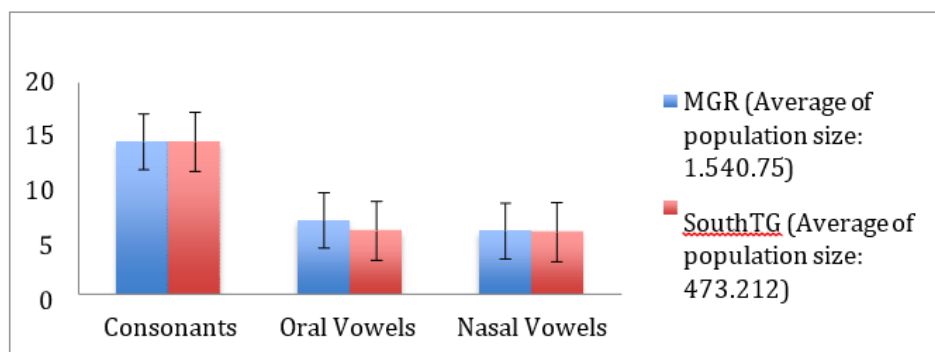
## Results and discussion

### Results

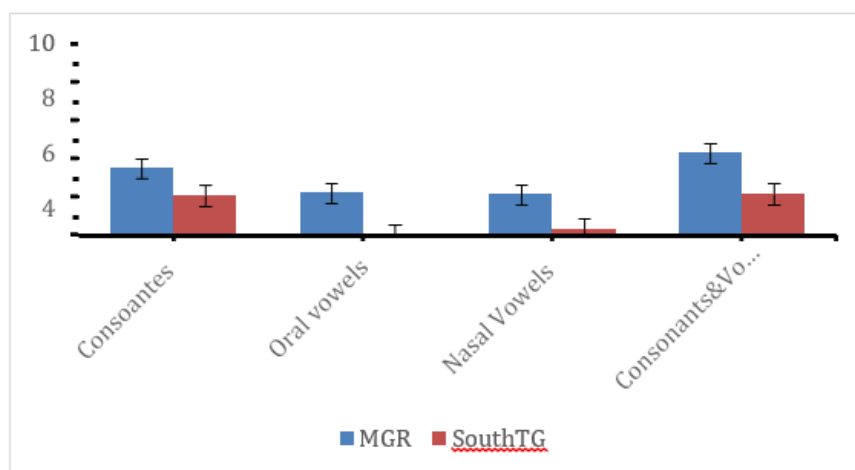
T-tests and tests of variance were performed in order to compare: (a) regions *vs.* number of consonants and number of vowels (oral and nasal), (b) populations size *vs.* number of consonants and number of vowels (oral and nasal). The T-test results were not significant. However, as figure 10 shows, there is a small difference in oral vowels: TupiMGR have more oral vowels than SouthTG.

In contrast, a variance test comparing TupiMGR with SouthTG was significant for vowels. There is more vowel-inventory-size variation (and consequently more vowel variation) among TupiMGR than among SouthTG, as shown in figure 11.

**Figure 10:** Region *vs.* Number of consonants, oral and nasal vowels



**Figure 11:** Variance test: Region *vs.* Number of Consonants, oral and nasal vowels and Consonants & vowels

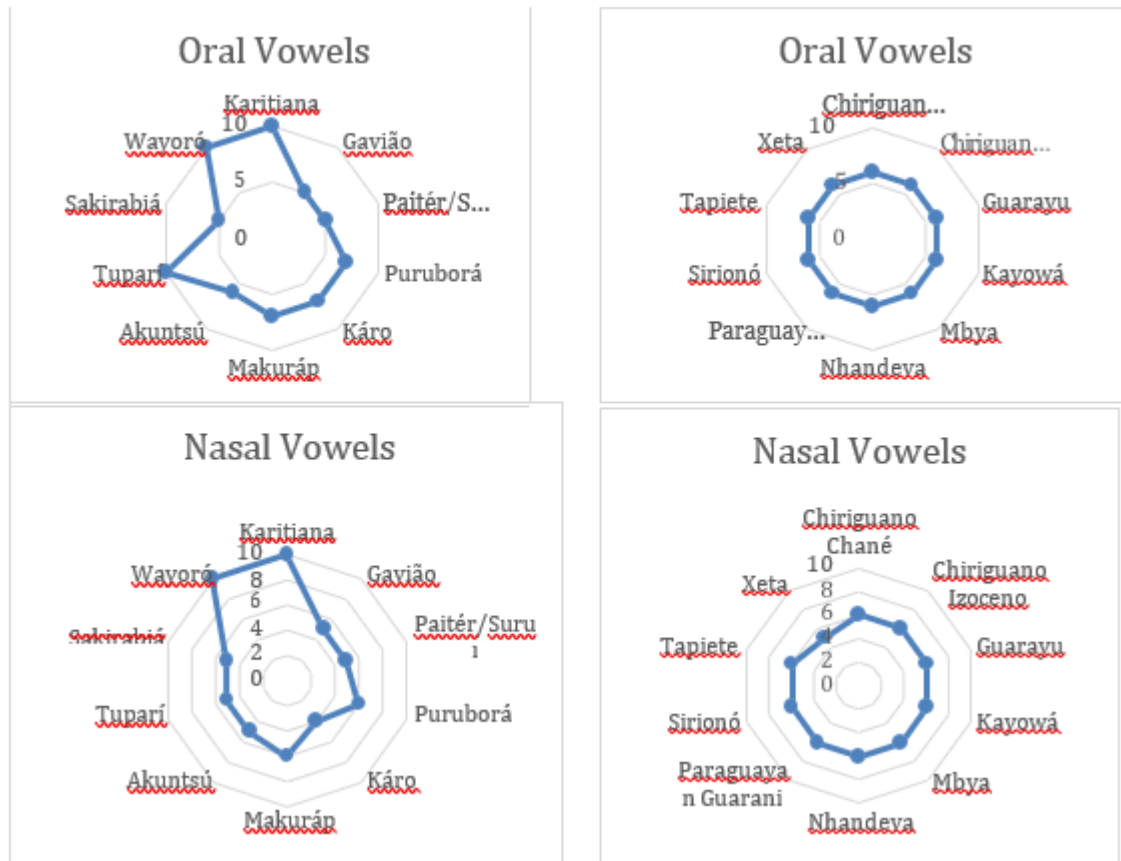


Oral Vowels  $P < 0.0001$

Nasal Vowels  $P < 0.0001$

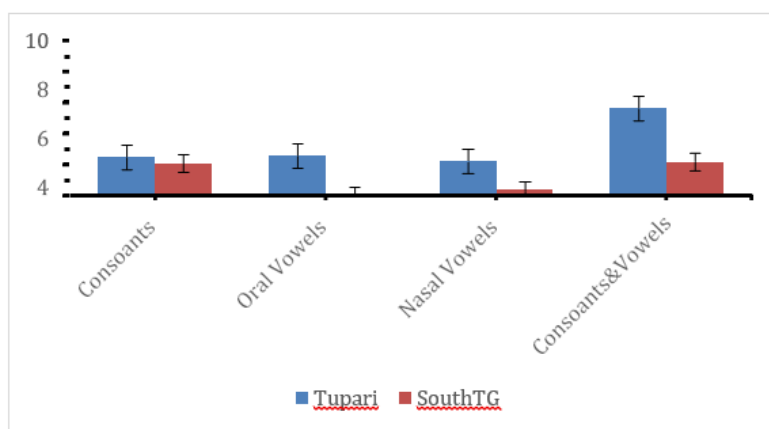
This result might be more apparent in the graphics below (figure 12):

Figure 12: Phonemic-inventory-size variation among TupiMGR and SouthTG



In the variance test above, all TupiMGR languages were clustered together. Hence, a multi-family factor might be affecting the results, as we are comparing a cluster of languages from different families (Tupari, Mondé, Arikém and Puruborá) with a group of languages from the same family (Tupí-Guarani). To verify this, we run another variance test comparing only Tupari languages versus SouthTG. The results were similar, shown in figure 13<sup>25</sup>.

Figure 13: Variance test: Language family (Tupari and Tupi-Guarani) vs. number of consonants, oral and nasal vowels and consonants & vowels



25 We could not run a variance test for the other MGR families because the samples for these families were too small.



Oral Vowels P < 0.0001

Nasal Vowels P < 0.0001

In sum, in comparison to TupiMGR, SouthTG exhibits a smaller vowel inventory, although the difference is not significant. In accordance, TupiMGR and SouthTG contrast significantly with respect to variation in vowel repertoire: TupiMGR presents more inter- language variation than SouthTG.

## Discussion

First, it should be observed that our sample was small. Thus, comparison between population size and phoneme inventory size may not have been significant for that reason. Also, as Cysouw et al. (2012) pointed out, population-size and phonemic-inventory-size interact statistically only when large populations (i.e. populations above 100.000) are considered (see section 3).

Comparative studies aiming at reconstructing Proto-Tupi-Guarani (ProtoTG) reaches a consensus with respect to vowels. ProtoTG has the same 6x2 harmonious series of vowels observed in modern SouthTG languages (6 oral, 6 nasal) (SCHLEICHER, 1998, MEIRA & DRUDE, 2015).

As for consonants, different inventory sizes have been assumed: 12 consonants in Lemle (1971), 13 Schleicher (1998), 19 in Rodrigues and Dietrich (1997)<sup>26</sup>. Thus, there are two possibilities:

- (a) ProtoTG had a smaller consonantal system, which was amplified by some SouthTG languages due to a population-size effect. (Paraguayan Guarani has the largest consonantal inventory among SouthTG languages (18 consonants), and it has by far the biggest estimated number of speakers (5.850.000))
- (b) ProtoTG had an inventory of 19 consonants, which was reduced in SouthTG due to a founder effect.

Given Perreault & Mathew (2012) observation that phonemes accumulate at rate between 0.26 and 0.38 per 1.000-year period (see section 3), possibility (a) is unlikely. Also, a statistical analysis done by Creanza et al. (2015), on the phoneme inventories of 2,082 worldwide languages, indicates that languages in contact present less variance in number of phonemes than isolated languages. Thus, the fact that Paraguayan Guarani coexists in a (quasi)-bilingual situation with Spanish suggests that this language might be conservative, being similar to Proto-TG with respect to number of consonants. Therefore, hypothesis (b) is favored<sup>27</sup>.

Coming back to vowels. In comparison with TupiMGR, SouthTG presents a non-

<sup>26</sup> See also Drude (2011) and Meira & Drude (2015).

<sup>27</sup> A reviewer observed that there might be a significant time difference between the increase in consonantal inventory and the contact with Spanish colonizers. Notice, however, that according to Perreault & Mathew, the phoneme accumulation rate is really small (0.26 to 0.38 per 1.000-year period).

significant smaller oral-vowel inventory, and a significant lack of inter-language variation in vocalic sounds (oral & nasal vowels). It is interesting that this contrast is placed on the vowel system, which according to Hay and Bauer (2007) are more resistant to fluctuations on population size. However, if ProtoTG already had a 6x2 vowel inventory, then SouthTG are just a cluster of conservative languages, preserving the vowels they inherited from ProtoTG.

In sum, if SouthTG experienced a reduction in their consonantal inventories (possibility (b)), but preserved the vowel system of ProtoTG, we conclude that SouthTG fits the general pattern of languages that underwent territorial expansions with fluctuation on population size (bottlenecks) and consequence founder effects<sup>28</sup>.

In addition, if Rodrigues (2007b) is right in postulating that Proto-Tupi had 28 consonants, Fort & Pérez-Losada's (2016) observation that the Proto-language(s) that left Africa already had a small phonemic inventory applies here as well because ProtoTG had a reduced phonemic inventory compared to Proto-Tupi.

Although this hasn't been discussed in the literature on phonemic founder effects, I believe our findings about Tupi converse well with the literature on the role played by consonant and vowels in I-language. It has been suggested that vowels and consonants do not have the same mental representation. Investigations on language acquisition and on phonemic perception indicate that there is a division of labor between vowels and consonants: while the task of consonants is identification and differentiation among lexical items, vowels are used to mark prosodic boundaries and structural relations. Vowels are signalers of syllabic structure, syntactic boundaries, and other syntactic properties such as constituent order (NESPOR et al., 2003; DONATTI et al., 2007; HOCHMANN et al., 2011). To appreciate this, one should consider that vowels, more frequently than consonants, are employed as markers of morphosyntactic processes. Take Guarani Mbyá as an example<sup>29</sup>. Most of the grammatical processes in Guarani Mbyá are morphologically realized by vowels. For instance, vowels are used as nominalizers:

- (1) a. ayvu -jopy -a (Martins, 2003: 123)  
*speech-grab-NMLZer*  
'recorder'
- b. Karu-a  
*to.eat-NMLZer*  
'place of eat'
- c. mba'e-xo-a  
*thing-pound-MNLZer*  
'pestle'

---

28 Xetá is the only SouthTG with 5 nasal vowels. It seems that Xetá behaves differently in grammatical features as well, which might be the result of language contact. See Michael et al. (2015) and reference therein.

29 I am pretty sure the same can be seen in better-known language like Brazilian Portuguese.

They are also the distinctive markers of agreement. As described on table 8, consonants are used on the composition of person agreement affixes. However, only vowels can occur alone (1PSg, 3PSg of active pronouns)<sup>30</sup> and contrastive distinctions are done by vowels (*oro* (active pronoun exclusive 1PPl) vs. *ere* (active pronoun 2PSg) and *ore* (inactive pronoun exclusive PPl)).

**Table 8:** Person agreement markers

Person	Active Pronouns	Inactive Pronouns
1a. p.sg.	a-	xe
1a. p.pl. (inclusive)	ja-	nhande
1a. p.pl (exclusive)	oro-(ro-)	ore
2a. p; sg.	ere-(re-)	nde (ne)
2a. p.pl.	pe-	pende (pene-)
3a. p.sg/pl	o-	

(MARTINS, 2003: 36)

If vowels are indeed structural markers, then it can be predicted that there is more structural cohesion (i.e., less parametric variation) among SouthTG languages than among MGR languages families. For example, Tuparí languages, as we saw above, present some inter-language degree of variation with respect to vowels.

We will not investigate this issue here, as it is beyond our present goals. However, it is important to observe that this prediction is in accordance with the fact that Tupi-Guarani languages are clustered in subgroups, with the language analyzed here belonging to two subgroups, as discussed above. We haven't yet looked at other Tupi-Guarani languages, but this an interesting way of tying down morphosyntax, phonology and phonetics, and it should be considered in our future analyses.

## Conclusion

Although genetic and languages are to be understood as different objects of evolution, a parallelism between genetic and phonemic information has been traced: territorial expansion may cause a population to reduce its genetic variability and the phonemic inventory of its language - *founder effects*. Investigations on American native people suggest that territorial dispersals within the continent have led to loss of genetic variability. Given this, we raised the question of whether a similar effect is observed within linguistics, with dispersed native populations presenting decays in their phonemic systems. To investigate this possibility, we considered the territorial dispersal of Tupi-Guarani people in the south and southern regions of South America. Taking for granted the hypothesis that Tupi emerged at the basins of Madeira and

<sup>30</sup> As a reviewer observed, this might be due to a parametric restriction on syllabic structure. However, the question is why such restriction is imposed in the first place. It might as well reflect the structural role played by vowels.

Guaporé Rivers in Amazon, we mapped modern Tupi-Guarani languages spoken in the south of Amazon (Argentina, Bolivia, Paraguay and South of Brazil) to verify whether these languages present less phonemic variation than Tupian languages spoken at the Madeira-Guaporé.

The results indicate a significant difference between these two sets of languages: In Madeira-Guaporé, there is more inter language variation with respect to vocalic inventory size. The South Tupi-Guarani languages are harmonious among themselves, presenting two symmetrical series of oral and nasal vowels (a 6x2 system). This suggests that there is less vowel variability among South Tupi-Guarani languages than among the Tupian languages spoken in Madeira-Guaporé. Although our results are preliminary because our database was small, they suggest that a founder effect might have occurred.

Putting together our findings and the general discussion on founder effects in language, we show that reconstruction analyses postulating that Proto-Tupi-Guarani had more phonemes than the modern Tupi-Guarani languages discussed here seem to be right. Also, the conservative vocalic system observed in Tupi-Guarani raises questions about parametric variation among the grammars of southwest Tupi-Guarani languages (groups I and II in Rodrigues' (1985) classification). If vowels are signalers of structural relations, it follows that these languages are structurally similar to each other.

Clearly, we need to enlarge our empirical database, including other Tupi languages in our sample, in order to explore these issues in depth. This is one of our near-future tasks.

We started the paper talking about language preservation, and we end it talking about how research on native people and their languages can contribute to a better understanding of how evolutionary forces, including random events (e.g. territorial expansion), can shape us genetically and linguistically. Thus, we hope that the facts presented above speak by themselves in highlighting the importance of preserving diversity.

## References

AGUILAR, A. C. *Contribuições para os estudos histórico-comparativos sobre a diversificação do sub-ramo vi da família linguística Tupi-Guarani*. Doctoral dissertation, Universidade de Brasília, 2015.

ALMEIDA, F.; NEVES, E. Evidências arqueológicas para a origem do Tupi-Guarani no Leste da Amazônia. *Mana – Estudos de Antropologia Social*, v. 21, 499-525, 2015.

ATKINSON, Q. Phonemic Diversity supports a founder effect model of language expansion from Africa. *Science*, v. 332, p. 346-349, 2011.

BARTON, N. H.; CHARLESWORTH, B. Genetic revolution, founder effects and speciation. *Annual review of ecology and systematics*, v. 15, p. 133-164, 1984.

BISSO-MACHADO, R., BORTOLINI, M. C., SALZANO, F. M. Uniparental genetic markers

*Diadorim*, Rio de Janeiro, vol. 22, especial (2020), p. 65-97, 2020.

in South Amerindians. *Genetic Molecular Biology*, v. 35, p. 365-387, 2012.

BRAGA, A. *A fonologia segmental e aspectos morfofonológicos da língua Makurap*. MA thesis, Universidade Estadual de Campinas, 1992.

BROCHADO, J. P. *An Ecological Model of the Spread of Pottery and Agriculture into eastern South America*. PhD dissertation. Urbana-Champaign: University of Illinois at Urbana-Champaign, 1984.

CAMPBELL, L. Areal linguistics. In Wright, J. D (ed.) *International encyclopedia of the social & Behavior sciences*. 2ed. P. 955-60, 2015.

CAVALLI-SFORZA, L. L. 2001. *Genes, people and language*. University of California Press.

CAVALLI-SFORZA, L. L.; MENOZZI, P.; PIAZZA, A. 1993. *The history and geography and human genes*. Princeton: Princeton University Press.

COOK, L. M.; GRANT, B. S.; SACCHERI, I. J.; MALLETT, J. Selective brid predation on the peppered moth: the last experiment of Michael Majerus. *Biology Letter*, v. 23, p. 602-612, 2012.

COSTA, C. *Apyngwa rupigwa: nasalização em Nhandewa-Guarani*. Doctoral dissertation, Universidade Estadual de Campinas, 2007.

CREANZA, N.; RUHLEN, M.; PEMBERTON, J. T.; ROSENBERG, N. A.; FELDMAN, M.W.; RAMACHANDRAN, S. A comparison of worldwide phonemic and genetic variation in human populations. *PNAS*, v. 112, p. 1265-1272, 2015.

CRYSTAL, D. *Language death*. Cambridge: Cambridge University Press, 2000.

CYSOUW, M.; DEDIU, D.; MORAN, S. 2012. Comments on “Phonemic diversity supports of serial founder effect model of language expansion from Africa”. *Science*, 335: 657. DIETRICH, W. O tronco Tupi e suas famílias de línguas. Classificação e esboço tipológico. Volker, N.; Dietrich, W. (eds.) *O Português e o Tupi no Brasil*. São Paulo: Editora Contexto, 2010, p. 9-25.

DLUGOSCH, K. M.; PARKER, I. M. Founding events in species invasions: genetic variation, adaptive evolution and the role of multiple introductions. *Molecular Ecology*, v. 17, p. 431-449, 2008.

DONATTI, L; PEÑA, M.; NESPOR, M.; MEHLER, J. 2007. On consonants and vowels, chickens and eggs. *Psychological Science*, v. 18, p. 924-925, 2007.

DRYER, M.; HASPELMATH, M. (eds.). 2005. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2005. (Available online at <http://wals.info>, Accessed on 2019-10-26.)

DUARTE, F. B. *Diversidade linguística no Brasil: a situação das línguas ameríndias*. *Calestroscópio*, v. 4, p. 27-62, 2016.

GABAS, N. J. Tupian Languages. In: BROWN, K. (ed.) *Encyclopedia of language and*

*Linguistics*. Amsterdam: Elsevier, 2006, p. 146-150, 2006.

GASPARINI, N. *Observations on sociolinguistiques et esquisse de la phonologie du siriono*. Langue Tupí-guarani de Bolivie. MA thesis, Lumière Lyon 2, 2012.

GONZÁLEZ, H. A. *A grammar of Tapiete (Tupi-Guarani)*. PhD dissertation. University of Pittsburgh, 2005.

GREENBERG, J. H. *The languages of Africa*. Bloomington: Indiana University Press, 1963.  
GÜLDEMANN, T. Clicks, genetics, and “proto-world” from a linguistic perspective. *University of Leipzig Papers on Africa, Languages and Literatures*, v. 29, 1-32. 2007.

FLEMING, L. Phoneme inventory size and the transition from monoplanar to dually patterned speech. *Journal of Language Evolution*, v. 2, p. 52-66, 2017.

FORT, J.; PERÉZ-LOSADA, J. Can a linguistic serial founder effect in Africa explain the worldwide phonemic cline? *Journal of the Royal Society Interface*, v. 13, 20160185, 1–9. 2016.

FUTUYMA, D. J. *Evolution*. Sunderland, Massachusetts: Sinauer Associates, INC, 2005.

HAY, J.; BAUER, L. Phoneme inventory size and population size. *Language*, v. 83, p. 388- 400, 2007.

HOCHMANN, J. R.; BENAVIDES-VARELA, S.; NESPOR, M.; MEHLER, J. 2011. Consonants and vowels: different roles in early language acquisition. *Developmental Science*, v. 14, p. 1467-1687, 2011.

HOCKETT, C. F. The Origin of Speech. *Scientific American*, v. 203, p. 88-111, 1960.

IVO, I. *Características fonéticas e fonológicas do Guarani no Brazil*. Doctoral dissertation, Universidade Estadual de Campinas, 2018.

KNEIP, A.; MELLO, A. Babel indígena: arqueologia e linguística pode ajudar a revelar origem e trajetória dos Tupis e Guaranis, que se dispersaram pelo território que se tornaria o Brasil. *Dossiê Nação Indígena*. 2013.

LATHRAP, D. W. *The upper Amazon*. New York: Praeger, 1970.

LEMLE, L. Internal classification of the Tupi-Guarani linguistic family. In.: BENDOR-SAMUEL, D. (ed.) *Tupi Studies*. Norman: Oklahoma: Summer Institute of Linguistics. 1971, p. 107-129.

LEWIS, M. P. *Ethnologue: Languages of the World*. Sixteenth edition. Dallas, Texas: SIL International, 2009. <http://www.ethnologue.com/16>.

MADDIESON, I.; PRECODA, K. *The UCLA Phonological Segment Inventory Database*, 1990. [http://web.phonetik.uni-frankfurt.de/upsid\\_info.html](http://web.phonetik.uni-frankfurt.de/upsid_info.html).

MAIA, M.; FRANCHETTO, B.; LEMLE, M.; DAMASO-VIEIRA, M. 2019. *Línguas indígenas e Gramática Universal*. São Paulo: Editora Contexto, 2019.

MARQUES, A. M.; RENESTO, E. Genetic variability and evidence of founder effect in *Hemiodus orthonops* (Characiformes: Hemiodontidae) from the upper Paraná River Basin, Brazil. *Acta Scientiarum, Biological Sciences*, V. 39, p. 53-58, 2017.

MAYR, E., Change of genetic environment and evolution. In.: HUXLEY, J.; HARDY, A. C.; FORD, E. B. (eds.) *Evolution as a Process*. London: Allen & Unwin. 1954, p. 157–180.

\_\_\_\_\_. *Animal species and evolution*. Cambridge, Massachusetts: Harvard University Press, 1963.

\_\_\_\_\_. *What evolution is*. New York, NY: Basic Books, 2013.

MCKUSICK, V. A. Ellis-van Creveld syndrome and the Amish. *Nature Genetics*, v. 24, p. 203-204, 2000.

MEIRA, S.; DRUDE, S. A summary reconstruction of Proto-mawetí-guarani segmental phonology. *Boletim do Museu Paranaense Emílio Goeldi. Ciências Humanas*, v. 10, p. 275- 69, 2015.

MELLO, A. A. S. *Estudo histórico da família linguística Tupi-Guarani: aspectos fonológicos e lexicais*. Doctoral dissertation, Universidade Federal de Santa Catarina, 2000.

MICHAEL, L.; CHOUSOU-POLYDOURI, N.; BARTOLOMEI, K.; DONNELLY, E; WAUTERS, V.; MEIRA, S.; O'HAGAN, Z. A Bayesian phylogenetic classification of Tupí-Guaraní. *Liames*, v. 15, p. 193-221, 2015.

MILLER, A. 2011. The representation of Clicks. In.: VAN OOSTENDORP, M. et al. (eds.) *The Blackwell Companion to Phonology*. New York: Wiley-Blackwell, 2011, p. 416-439.  
MARÇOLI, O. *Estudos comparativos dos dialetos da língua Kawahib (Tupi-Guarani): Tenharim, Jiahui e Amondawa*. MA thesis, Universidade Estadual de Campinas, 2018.

MARTINS, M. F. *Descrição e análise de aspectos da gramática do Guarani Mbyá*. PhD. Dissertation. Universidade Estadual de Campinas, 2003.

MOORE, D.; MEYER, J. The study of tone and related phenomena in an Amazonian tone language, Gavião of Rondônia. *Language Documentation & Conservation*, v.8, p. 613-636, 2014.

NESPOR, M.; PEÑA, M.; MEHLER, J. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, v. 2, p. 2003-2229, 2003.

NETTLE, D.; ROMAINE, S. *Vanishing voices: the extinction of the world's language*. Oxford: Oxford University Press, 2000.

NIMUENDAJÚ, C. The Cawahib, Parintintin and their neighbors. In: STEWARD, J. (ed.) *Handbook of South American Indians, Vol. 3*. Washington: Government Printing Office. 1948, 283-297.

NOELLI, F. As hipóteses sobre os centros de origem e as rotas de expansão dos tupi. *Revista de Antropologia*, v. 39, p.7-53, 1996.

PAYNE, M.; RUPAR, A.; SIU, G. M.; SIU, V. M. Amish, Mennonite and Hutterite genetic disorder database. *Paediatric Child Health*, v. 16, p. e23-e24, 2011.

PERREAULT, C.; MATHEW, S. Dating the Origin of Language Using Phonemic Diversity. *PLoS One*, v. 7, p. e352-89, 2012.

PIERCE, S. B.; YOST, C.; BRITTON, J. S.; LOO, L. W. M.; FLYNN, E. M.; EDGAR, B.A.; EISENMAN, R. N. dMyc is required for larval growth and endoreplication in *Drosophila*. *Development*, v. 131, p. 2317-2317, 2004.

RAMALLO, V.; BISSO-MACHADO, R.; BRAVI, C.; COBLE, M. D.; SALZANO, F. M.; HUNEMEIER, T.; BORTOLINI, M. C. Demographic Expansions in South America: Enlightening a Complex Scenario with Genetic and Linguistic Data. *American Journal of Physical Anthropology*, v. 150, p. 453-463, 2013.

RAMACHANDRAN, S.; DESHPANDE, O.; ROSEMAN, C.; ROSENBERG, N.; FELDMAN, M.; CAVALLI-SFORZA, L. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, v. 102, p. 15942–15947, 2015.

RIDDLEY, M. *Evolution*. Third edition. Backwell, 2004.

RITO, T.; VIEIRA, D.; SILVA, M.; CONDE-SOUSA, E.; PEREIRA, L.; MELLARS, P.; RICHARDS, M.; SOARES, P. A dispersal of *Homo sapiens* from southern to eastern Africa immediately preceded the out-of-Africa migration. *Scientific Reports*, 9, 2019.

RODRIGUES, A. D. 1964. Classificação do tronco linguístico Tupi. *Revista de Antropologia*, v. 12, p. 99-104, 1964.

\_\_\_\_\_. Relações internas na família Tupí-Guaraní. *Revista de Antropologia*, v. 27, p. 33- 53, 1985.

\_\_\_\_\_. *Línguas brasileiras: para o conhecimento das línguas indígenas*. São Paulo: Loyola, 1986

\_\_\_\_\_. 2005. As vogais orais do Proto-Tupí. In.: RODRIGUES, A. D.; CABRAL, A. S. A.C. (eds.) *Novos estudos sobre línguas indígenas*. Brasília: Editora da UnB, 2005, p. 35-46.

\_\_\_\_\_. Tupí languages in Rondônia and in eastern Bolivia. In.: WETZELS, L. (ed.) *Language endangerment and endangered languages*. Leiden: CNWS Publications. 2007a, 355-363.

\_\_\_\_\_. As consoantes do Proto-Tupí. In.: CABRAL, A. S. A. C.; RODRIGUES, A. D. (eds.) *Línguas e culturas Tupí*. Campinas: Editora Curt Nimuendajú, 2007b, p.167-203.

\_\_\_\_\_. *Línguas indígenas brasileiras*. Brasília. LALI - Laboratório de Línguas Indígenas da UnB, 2013. <http://www.laliunb.com.br>

\_\_\_\_\_.; DIETRICH, W. 1997. On the linguistic relationship between Mawé and Tupi- Guarani. *Diachronica*, v. 14, p. 265-304. 1997.



\_\_\_\_\_; CABRAL, A. S. A. C. Revendo a classificação interna da família linguística Tupí-Guaraní. In.: CABRAL, A. S. A. C.; RODRIGUES, A. D. (eds.) *Línguas Indígenas Brasileiras: Fonologia, gramática e história*. Belém: UFPA, 2002, p. 327-337.

\_\_\_\_\_.; \_\_\_\_\_. 2012. Tupían. In.: CAMPBELL, L.; GRONDONA, V. (eds.). *The indigenous languages of South America*. Berlin: Mouton de Gruyter, v. 2, 2012, p. 495-574. SAMPAIO, W. *Estudo comparativo sincrônico entre o Parintintin (Tenharim) e o Uru-eu- uau-uau (Amondwa): contribuições para uma revisão na classificação das línguasupi-Kawahib*. MA thesis, Universidade Estadual de Campinas, 1998.

SANTOS, E.; SILVA, A.; EWERTON, P.; TAKESHITA, L.; MAIA, M. Origins and demographic dynamics of Tupí expansion: a genetic tale. *Boletim do Museu Paranaense Emílio Goeldi. Ciências Humanas*, v. 10, p. 217-28, 2015.

SCHLEICHER, C. *Comparative and internal reconstruction of the Tupi-Guarani language family*. PhD Dissertation. University of Wisconsin, Madison. 1998.

SIMONS, G. F. Two centuries of spreading languages loss. *Proceedings of the Linguistic Society of America*, v. 4, p. 1-12. 2019.

SPROAT, R. Phonemic diversity and the out-of-Africa theory. *Linguistic Typology*, v. 15, p. 199–206, 2011.

STORTO, L.; DEMOLIN, D. *Pitch accent in Karitiana*. Ms., Universidade de São Paulo, 2005.

TRUDGILL, P. Social Structure and Phoneme Inventories. *Linguistic Typology*, v. 15, p. 155-60, 2011.

URBAN, G. A história da cultura brasileira segundo as línguas nativas. In.: CUNHA, M. C (ed.) *História dos índios no Brasil*. São Paulo: Companhia das Letras. 1992, p. 87-102.

URBAN, G. On the geographical origins and dispersion of Tupian languages. *Revista de Antropologia*, v. 39 p. 61-104, 1996.

VASCONCELOS, E. *Aspectos fonológicos das línguas Xetá*. MA thesis. Universidade de Brasília, 2008.

YANG, N.N.; MAZIÈRES, S.; BRAVI, C.; RAY, N.; WANG, S.; BURLEY, M. W.; BEDOYA, G.; ROJAS, W.; PARRA, M.V.; MOLINA, J. A. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Annals Human Genetics*, v. 74, p. 525–538, 2010.