DOMAINS AND PROMINENCE IN NASAL HARMONIZATION OF MAXAKALÍ LOANWORDS[1*]

MÁRIO COELHO DA SILVA, ANDREW NEVINS AND JAMES WHITE

1. Overview

In the literature on loanword phonology, there is a longstanding debate as to whether the adaptation from the source language is based on phonetic similarity from language-to-language or mediated by underlying representations (e.g. Silverman 1992, Kenstowicz 2003, Peperkamp and Dupoux 2003, LaCharité and Paradis 2005). As originally pointed out in Wetzels (2009), nasality in Maxakalí, a Macro-Jê language spoken in Minas Gerais, Brazil, turns out to be an important test case in this debate, as there is a strong effect of constraints banning a mismatch in nasality between consonant and vowel, e.g. *[dã] and *[ma], potentially affecting the adaptation of numerous loanwords. For instance, in order to incorporate a Brazilian Portuguese (BP) loanword like *martelo* [mahˈtɛlʊ]'hammer' into the language, the first syllable must be turned into either fully nasal [mã] or fully oral [ba] in Maxakalí. In the native phonology of Maxakalí, consonant nasality is argued to be determined by neighboring vowels, rather than vice-versa (Wetzels, 2009). However, as it turns out, the preferred choice in such loanwords is often [mã]: the vowel 'loses' in terms of surface similarity, so that the nasality of the consonant may be preserved. This finding is surprising, in that nasality is introduced onto what were contrastively oral vowels, but on the other hand, it may reflect a loanword adaptation bias to preserve the nasality of initial consonants from source words. It is precisely this negotiation between conflicting forces that a model of loanword phonology must address.

In the present paper, we contribute to an understanding of the patterns of loanword adaptation (and therefore the representation of nasality in word-level phonotactics in Maxakalí

more broadly) through the inclusion of data from 768 word tokens (95 types) in which constraints on nasality must be negotiated in the process of loanword phonology. In doing so, we separate out the application of nasal harmony in the language into three potential domains:[2]

(1) Potential Domains of nasalization in Maxakalí word-level phonology:

-*Rime Nasalization* Within a rime, the vowel and any consonantal coda will agree in nasality.

-*Onset Nasalization* Within a syllable, the onset consonant, if it is not voiceless, will agree in nasality with the vowel.

-*Intersyllabic Harmony* All vowels to the left of the stressed vowel (the language has consistent final stress) may pick up the nasality of the stressed vowel, but can be blocked if a voiceless consonant intervenes

As for Onset Nasalization, this is reflected in the writing system, whereby <mã> indicates [mã] and <ma> indicates [ba], specifically emphasizing the vowel as a determinant of the nasality of a voiced onset consonant). Regarding intersyllabic harmony, this would be as a nucleus-to-nucleus relation, as proposed in Piggott & van der Hulst 1997 for languages such as Barasano. Many models of nasal harmony (and of word-level nasalization in languages of this sort) conflate all three kinds of targets into a single process. As native words such as (2) and (3) in fact show nasalization of every segment in the word, it can be difficult to tease apart precisely which nasalization processes are at play in Maxakalí based on the native vocabulary.

(2) [ɲãmĩɲ] ‘spirit’

(3) [ɲĩmãŋ] ‘wing’

Gudschinsky, Popovich & Popovich (1970) proposed that both nasal vowels and nasal consonants are phonemic, with a rule of consonant denasalization before oral vowels (essentially [-nasal] harmony). On the other hand, Rodrigues (1981) proposed that, given that voiced stops are always nasal in word-final codas in Maxakalí, only voiced stops are underlying, and the distribution of nasality results from the predictable rule of word-final nasalization plus leftward nasal harmony. In Rodrigues’ proposal, voiceless consonants block nasality spreading, so a word such as [kuˈkũñ̪], *[kũˈkũñ̪], ‘to clear one’s throat’ will come from an underlying form /kokod/, in which the voiced consonant nasalizes and subsequently spreads its nasality leftwards up until the voiceless consonant that stops further leftward spreading.[3] Finally, approaches such as Walker (2003) take nasal harmony to be bidirectional for languages such as Tuyuca, and under this style of analysis, nasalization would spread from the stressed vowel to every other segment in both directions, except when halted by voiceless consonants.

In contrast to monolithic approaches to nasal harmony in terms of a single underlying source for nasality that either spreads bidirectionally or unidirectionally, a number of researchers have posited that the processes of syllable-internal nasalization (herein called ‘nasal concord’, or syllable-level phonotactics) and syllable-to-syllable nasalization (‘harmony’) are distinct (Piggott & van der Hulst 1997, Wetzels 2009, Thomas 2014, Singerman 2016). This cleaving of harmony into two (or more) processes would suggest that the processes in (1) could ‘go their own way’ when speakers are confronted with novel phonetic sequences to be adapted from donor languages, thereby providing the opportunity to produce novel combinations according to their phonology.

Such a state of affairs – whereby a set of phonotactic constraints that are generally

exceptionless in the native phonology nonetheless break down in different ways and at different rates in the adaptation of loanwords – has been found before in studies of loanword phonology, leading to what is called *differential importation* (Kang 2011). As Broselow et al. (2012) describe, for example, "Although neither [ti] nor [si] is attested in native Japanese vocabulary, [ti] is more likely than [si] to be preserved in loanwords, as illustrated by the pronunciation of 'Citibank' as [ʃitibaŋku], where [s] becomes palatalized but [t] is retained".[4] They analyze this asymmetric importation pattern as reflecting a native language ranking *[SI] >> *[TI], under which these two distinct phonotactic constraints, though they go unviolated in the native language, reveal their relative importance via loanword phonology. Similarly, Smith (2018) documents what are called 'impossible nativization effects', such that if segmental or phonotactic phenomenon *x* is adapted in a loanword, then *y* must be adapted as well (i.e. *y* is not possible without also having *x*). In the present paper, we investigate a logic of comparing rates of adaptation with respect to the three domains of nasalization presented in (1) in Maxakalí loanwords from Portuguese, investigating whether the 'harmonization', i.e. imposition of nasality through harmony, is more likely to occur in particular domains within the word. As will be shown below, Maxakalí speakers enforce nasal harmony more strongly within smaller domains. In addition, they are more likely to preserve orality/nasality contrasts from BP in prominent positions, such as initial syllables and stressed syllables.

1.1 Phonological Overviews of Maxakalí and Brazilian Portuguese

Maxakalí has 1,907 speakers distributed across four villages around the Mucuri river valley, in the northeast of the state of Minas Gerais, Brazil, according to the 2013 SIASI census. Despite over 300 years of contact with Portuguese speakers, they still strongly maintain their cosmology, ritual, and the use of their first language, with a fairly high degree of monolingualism. The

group's autonym is [tɪjcmɨ̃ˈʔɨ̃ɰ̃ə̃], and they call their language [ɨ̃ɲĩ̃ˈʔɑj]. The language is ergative and has primarily Subject Object Verb word order (see Campos 2009 for an overview of its syntax and morphology), with largely an isolating profile in morphology, though with extensive compounding. Maxakalí phonology is well-known for its stop-consonant vocalization process, whereby obstruent coda consonants can lenite to glides (Gudschinsky, Popovich & Popovich 1970, Silva 2015). /p, t, c, k/ lenite to [β̞, ə̯, j, ɰ] respectively, while nasal codas lenite to the nasal version of the aforementioned glides. Words in Maxakalí generally follow a CVCVC template, with a final closed syllable (and even, in some older loans, the epenthetic insertion of a final consonant to satisfy this template; Araújo 2000, Ribeiro 2012). Thus, BP loanwords typically lose their final vowel, e.g. *to.ma.te* > [to.mãə̯] 'tomato' (alongside gaining nasality and lenition of final C). This truncation of the final unstressed vowel usually results in alignment of the stressed syllable of the source with the stressed syllable of the loan, and it causes BP onset consonants to become codas. In Maxakalí, codas are found only word-finally in morphologically simplex words, and can be drawn from any of the consonants in the language except /h/.

The consonantal inventory includes four voiceless stops, /p, t, c, k/ (represented in the standard orthography, developed by Popovich & Popovich (2005), as <p,t,x,k>), which resist nasalization and in fact block harmony further to the left. The palatal voiceless obstruent has affricate, fricative, and stop realizations. The language has a set of four pairs of obstruents, which alternate between voiced and nasal stops: [b~m, d~n, ɟ~ɲ, g~ŋ]. Each pair is represented by a single grapheme in the orthography (<m,n,y,g>, respectively), and within the native phonology, due to the processes of Rime Nasalization and Onset Nasalization in (1), their allophonic occurrence is fully predictable depending on the nasality of the syllable nucleus; thus [b] occurs before oral [a] while [m] occurs before nasal [ã]. However, the velar nasal is banned from onset position, and only found in codas, and thus the oral allophone /g/ is found in onset position

whether or not the following vowel is nasal (cf. *gõy* /gõɟ/ ‘smoke’ vs *gox* /goc/ ‘ambush’) In addition, there are two laryngeal consonants, /ʔ, h/, whose distribution is rather limited, with /ʔ/ occurring at morpheme boundaries (which otherwise would begin/end with vowels) and between vowels, and /h/ that is only found in onset position (except for some loanwords). Maxakalí lacks many consonants present in Portuguese (/l/, /r/, /f/, etc.; the name of the language, [maʃakaˈli], is in fact an exonym[5]). The vocalic phonemes of the language contrast five oral vowels /i, e, a, ɨ, o/ and their nasal counterparts /ĩ, ẽ, ã, ɨ̃, õ/.[6]

Brazilian Portuguese, by contrast, has a different phonological structure. Portuguese has complex onsets (largely obstruent-liquid clusters), a range of sibilant fricatives, three nasal consonants (m, n, ɲ) alongside voiced stops (b, d, g), five nasal vowels that can be lexically contrastive, regressive allophonic nasalization of vowels by adjacent consonants, and a restriction of coda consonants to sonorants and sibilants. Stress in nouns is lexical, falling within a three-syllable window. Crucially, outside of regressive vowel nasalization caused by adjacent nasal consonants (e.g. [kə̃.mə] ‘bed’), there is no other process of nasal harmonization in Portuguese affecting vowels or consonants. Brazilian Portuguese does not have any phonotactic requirement that onsets match nasality with tautosyllabic nasal vowels. For a broader overview of Portuguese phonology, see Mateus & D’Andrade (2000) and Cristófaro Silva (2007).

According to Rodrigues (1981), nasal harmony is right-to-left in Maxakalí, starting with the final segment. He posits a rule whereby morphemes ending in a voiced oral consonant undergo nasalization and another rule that subsequently enforces the spreading of its nasality to voiced segments to its left. Thus, /ɟabiɟ/ becomes [ɲãˈmĩɲ] ‘spirit’ due to an imperative to undergo predictable nasalization at the right edge (as words do not end with voiced stops) and then spread this nasality all the way to the left within the word (halted by voiceless stops). However, cases such as [ɲãˈmĩɲ] can be accounted for by any model of harmony: left-alignment,

right-alignment, pairwise agreement, syllable-bound harmony, under many combinations of assumptions of how much nasality is underlying (including whether it is underlyingly contrastive only on consonants, only on vowels, or both). All of these models are *consistent with data in which every segment shows surface agreement in nasalization.*

Thus, in Rodrigues' (1981) account, harmony is seen as a kind of left alignment of nasality emanating from a single source, where one single specification and source of nasality is associated and aligned as far as possible to the left edge. However, it is also possible to view nasal harmony as a pairwise, *local* harmony relation, in which each adjacent pair of voiced segments must agree in nasality. The strength of each such harmony relation may be enforced by factors such as syllabic constituency, positional factors, and segmental properties. Under such an account, there can be multiple possible sources of nasality per word. Thus, particularly in loanwords from Portuguese where multiple segments may have nasality in the source form, there are multiple sources of nasality that could enforce harmony in neighboring segments.

Wetzels (2009), working from dictionary data (primarily Popovich & Popovich 2005), makes a number of important observations that form the basis for this article. He observes that Rime Nasalization is exceptionless in Maxakalí loanwords from Portuguese (e.g. *tomate* [tuˈmaʧi] > [tũˈmãʒ̃] 'tomato'). Unlike Rodrigues (1981), however, who predicts that voiced codas should always become nasal in Maxakalí, Wetzels (2009) allows for oral adaptation of loans from Portuguese, such as *soldado* > [ʦuˈdɑɔ̃] 'soldier'. According to Wetzels, what governs the choice of whether a rime will be nasal or not also involves the ability to license the nasal onset. Thus, in *tomate* > [tũˈmãʒ̃] 'tomato', nasality in the rime will in turn license nasality in the onset of the second syllable, thereby allowing a faithful mapping of this consonant from Portuguese. Nonetheless, what Wetzels and Rodrigues have in common is the prediction that nucleus and coda must agree in orality/nasality in Maxakalí. This exceptionless character already

suggests a different status for Rime Nasalization, which is to be contrasted with the two other processes discussed below. Importantly, Rime Nasalization as found in Maxakalí is not inherited directly from Brazilian Portuguese, as BP has forms like [hɐ̃s] 'frogs' or [tɾɐ̃s.ˈpɔh.ʧi] 'transport', in which a nasal vowel is followed by an oral consonant.

Wetzels notes that unlike Rime Nasalization, Onset Nasalization is variable (e.g. he reports (4)) and even finds denasalization of the onset (e.g. (5)).

(4) *feijão* [feˈʒə̃ʊ̃] > [pɛ̃ˈn̲õŋ̲] or [pɛ̃ˈʒõŋ̲] 'beans'[7]

(5) *limão* [liˈmə̃ʊ̃]> [di̲bã̲β̃] 'lime'

These latter patterns suggest the potential future emergence of a /b~m/ contrast in Maxakalí as unpredictable, rather than allophonic, as anticipated in Wetzels (2009). We aim to investigate the extent of this variability. Finally, we sought to study in greater detail the rate of vowel-to-vowel nasal harmony, given that it can be modeled as a distinct phonological process. Our empirical results below, therefore, are designed to compare the amount of adherence to nasalization in the three domains in (1) above, with the hypothesis that the smaller the domain, the more adherence there will be. In other words, we set out to test whether there would be a higher percentage of rime nasalization than onset nasalization, and in turn a higher percentage of onset nasalization than intersyllabic harmony. The existence of reliable differences in the rates of application of these three processes would in turn suggest that they should be treated as distinct processes in terms of the phonological grammar of the language. This kind of variability is ideally investigated with a broad range of speakers, which is what motivated our data collection, explained shortly below.

1.2 Accounting for Variability and Exceptionful Generalizations

As discussed above, bringing BP loanwords into Maxakalí forces a clash between two phonologies: that of BP, in which nasality on consonants is fully phonemic (and allophonically predictable on stressed vowels), and that of Maxakalí, in which consonant nasality is mostly predictable. The adaptation strategies involved, however, have already been noted as variable, including a gain of nasality on vowels, a loss of nasality on consonants, or preservation of disharmonic sequences. The reason for such variability may turn out to result from multiple factors, including both demographic factors (e.g. speaker age or gender) and phonological factors (e.g. the position of the nasals within the word or the place and manner of articulation of the consonants involved). In order to address these questions with a more precise quantitative picture of the trends involved, we collected loanword data from 18 speakers (described in Section 2), analyzed the trends observed in the collected corpus (Section 3), and finally implemented a maximum entropy (MaxEnt) phonological model (Goldwater & Johnson 2003) to assess the relative weights of each factor.

The MaxEnt model, presented in Section 4 provides insight into the relative strength of the multiple potential generalizations that hold of the data. Anticipating the results, Maxakalí speakers do enforce nasal harmony, with stronger enforcement within smaller domains, and they tend to preserve orality/nasality contrasts from BP in prominent positions such as initial syllables and stressed syllables.

2. Materials and Data Collection

We conducted an analysis of loanword adaptation with 18 speakers of Maxakalí, balanced across three age groups (youngsters, age 15-29; adults, age 30-44; elders, age 45 and up) and two genders (female and male). The age groups roughly correspond to the span of a generation in the community in terms of parents, grandparents, and great-grandparents. In addition to providing a broader empirical base on which to study the extent of variation in loanword adaptation, it allows us to consider whether demographic factors played a role. This followed the design of prior work on Maxakalí within the domain of consonant lenition (Silva & Nevins 2014) that examined demographic factors and found a change-in-progress, whereby the elders produced consonant lenition much less than adults and youngsters. The current data were collected in *Apne Yĩxux* ('the Green Village'). Speakers were auditorally presented with a word in Portuguese (as pronounced by a native speaker from Belo Horizonte, Brazil) known to have an existing loanword variant in Maxakalí, and were then asked for the loanword's pronunciation. We elicited an average of 43 loanwords from each of the 18 speakers, with a total of 95 distinct loanword types, yielding an overall total of 768 word tokens. Those loanwords were selected based on their mention in previous research on Maxakalí, including Gudschinsky, Popovich and Popovich (1970), Araújo (2000), and Wetzels (2009). In some cases, we asked speakers to give us words in the same semantic field as other ones. We note that not every speaker produced a loanword adaptation for every form (e.g. some produced fully native neologisms, such as *kõnãg ãxi-nã ax* (water cool-CAUS NMZ) for 'fridge', *pip xo-nix* (metal point-two) for 'scissors' and *ã-yĩy ax ã-xape pu* (2sg-talk NMZ 2sg-relative DAT) 'cellphone'). For phonological adaptations where the speaker knew the loanword in Maxakalí, coding of each segment as nasal or oral was

performed by the first author, a native speaker of BP with extensive practical experience in monolingual Maxakalí communities.

As an overview, we found a wide range of adaptation strategies (as highlighted by underlining below):

(6) a. Gain of nasality on vowel: BP *martelo* [ma̲hˈtɛlʊ] > Mxk [mã̲hˈtɛ̰ə̰] ‘hammer’

b. Loss of nasality on consonant: BP *canivete* [ka̲ni̲ˈvɛʧɪ] > Mxk [kɨ̲di̲ˈbɛə̰] ‘pocket-knife’

c. Gain of nasality on consonant: BP *ta devendo* [tadeˈvẽnʊ] > Mxk [tadɛˈm̲ɛ̃ə̰] ‘to owe’

d. Maintenance of disagreeing CṼ: BP *Belo Horizonte* [bɛloriˈzõ̲ʧɪ] > Mxk [bɛdadiˈʤũ̲ə̰] ‘Belo Horizonte’

e. Variability in extent of harmony: BP *flamengo* [flaˈmẽgʊ] > Mxk [pã̲nã̲ˈmæ̃j̃], [pa̲nã̲ˈmæ̃j̃], [pa̲dã̲ˈmæ̃j̃] ‘flamengo (soccer team)’

f. Spontaneous nasalization (without an obvious source): BP *Oliveira* [o̲liˈverə] > Mxk [ũ̲nĩ̲ˈbɛə̰] ‘proper name’; BP *açúcar* [a̲ˈsukəh] > Mxk [ã̲ˈʧok] ‘sugar’

We report on these trends in quantitative and graphical format in the next section, before turning to the MaxEnt analysis.

3. Overview of corpus patterns

In this section, we provide an overview of some of the main patterns found in the corpus of loanwords. Given the extensive MaxEnt analysis in section 4, we do not provide an exhaustive statistical analysis for every possible comparison in this section. However, we report Fisher’s exact tests for particular comparisons of interest.[8]

3.1. Nasality agreement by domain

First, we consider the strength of the nasality agreement restriction within different prosodic domains. Figure 1 shows the proportion of adjacent consonant-vowel pairs (voiced sounds only) that agree in nasality in Maxakalí within three domains—VC within a rhyme, CV within a syllable, and V.C across syllables—separated according to whether they agree or disagree in nasality in the Portuguese source form.

The first observation that we can make is that nasality agreement is absolute within the rhyme, regardless of whether there was agreement in the Portuguese source word, whereas agreement outside if the rhyme is not absolute.[9] Second, adjacent sounds that already agree in Portuguese usually continue to agree in Maxakalí in all three domains; however, for adjacent sounds that disagree in nasality in Portuguese, there is often a change in nasality such that they agree in Maxakalí. The likelihood of agreement depends on the domain: a vowel and coda within a rhyme always agree (100%; e.g. Mxk [ˈbə̃ʉ̃ɻ]‘bank’ < BP [ˈbə̃kʊ]), a vowel and onset within a syllable agree somewhat less often (76%; e.g. Mxk [kãmãˈdoɯɻ] ‘horse’, < BP [kaˈvalʊ]), and an adjacent vowel and onset in separate syllables agree even less often (42%; e.g. Mxk [ʤĩ.nẽə̃] ‘window’ with agreement < BP [ʒaˈnɛlə], vs. Mxk [ɪjtaʤu.nĩə̃] ‘USA’ without agreement < BP [istadzuˈnidʊs]; the appendix contains a full list of adaptations, including interspeaker variability, which includes cases of VC (dis)agreement). Fisher’s exact tests support this stepladder effect, showing a significant difference between the VC group and the CV group ($p < .001$) and a significant difference between the CV group and the V.C group ($p < .001$). Thus, we see a hierarchy of domains when it comes to the strength of the nasality agreement restriction:

within rhyme > within syllable > across syllables (where > denotes having a stronger agreement restriction).

[INSERT FIGURE 1 HERE]

3.2. Nasality by position within the word

Next, we consider how position within the word affects the likelihood that onsets and vowels will remain faithful in nasality to the Portuguese source word. We set aside codas here because, as we have already seen, codas always agree in nasality with the nucleus in Maxakalí. Figure 2 shows the proportion of vowels and onset consonants that remain faithful in nasality, separated by syllable position within the word (initial, medial, or final) and the nasality of the sound in Portuguese. Monosyllabic words were included in both the initial and final groups in the plot.

First, we see that word-initial onsets are always faithful in nasality to the Portuguese source word, regardless of whether they are oral or nasal (e.g. Mxk [b̲ɤ̃ŋ] ‘bank’ < BP [ˈb̲ãkʊ]; Mxk [m̲ũˈʧa] ‘girl’ < BP [ˈm̲osə]). This absolute pattern suggests a strong effect of word-initial faithfulness. Second, we see that nasality is more likely to be preserved than orality, especially in vowels. In all syllable positions, nasal vowels are more faithful than oral vowels (initial, $p < .001$; medial, $p = .006$; final, $p < .001$). In non-initial syllables, nasal vowels in Portuguese in fact always remain nasal in Maxakalí in our sample (e.g. Mxk [nũˈb̲ɛ̃β̃] ‘November’ < BP [noˈv̲ẽbɾʊ])

whereas oral vowels sometimes become nasal (e.g. Mxk [hɛʰˈmɛ̃ɛ̯̃] ~ [hɛ̃ˈmɛ̃ɛ̯̃] ‘medicine’ < BP [hɛˈmɛdʒʊ]).

[INSERT FIGURE 2 HERE]

In initial syllables, there is a notable decline in faithfulness for vowels compared to medial and final syllables, which holds for both nasal and oral vowels (initial vs. medial nasal vowels, $p = .005$; all others, $p < .001$). However, a closer look at initial syllables suggests that the pattern is more nuanced in this position. Figure 3 shows only vowels in initial syllables according to the type of onset and the nasality of the source vowel in Portuguese. We see that nasal vowels in the initial syllable of Portuguese words always retain their nasality in our sample, unless they are preceded by a voiceless onset, where they sometimes become oral (e.g. Mxk [kuputaˈdo] ‘computer’ < BP [kõputaˈdoh]).[10] Strikingly, oral vowels in the initial syllable often spontaneously become nasal in Maxakalí even in the absence of any onset (e.g. Mxk [ãˈgʊj] ‘August’ < BP [aˈgostʊ]), though they usually remain oral when preceded by a voiced oral consonant. When preceded by a nasal onset, oral vowels in the initial syllable become nasal without exception in our sample (e.g. Mxk [mɛ̃ˈd̻z̻a] ‘table’ < BP [ˈmezə]).

[INSERT FIGURE 3 HERE]

Figure 4 takes a closer look at initial-syllable vowels preceded by voiceless consonants. We see a clear difference depending on whether the vowel is also followed by a voiceless consonant (meaning it is flanked on both sides by voiceless consonants) and whether it is followed by some other consonant. If the vowel is flanked on both sides by voiceless consonants, there is a strong

tendency for it to be oral in Maxakalí, regardless of whether it was oral or nasal in Portuguese (e.g. Mxk [kuˈpɑ̯ɐ̯] ‘compadre’ < BP [kũˈpadɾɪ] vs. Mxk [kãˈgɑj] ‘yoke’ < BP [kɐ̃ˈgaʎə]). The same preference does not hold when some other consonant follows ($p < .001$, for both oral and nasal vowels).

Taken together, these observations suggests that several pressures influence the likelihood that sounds will be realized as nasal in initial syllables, in addition to the general preference for adjacent sounds to agree in nasality: a strong preference for faithfulness in word-initial sounds (especially consonants), an overall preference for nasality in the initial syllable (especially for vowels), and a preference for oral vowels to occur between two voiceless consonants.

[INSERT FIGURE 4 HERE]

3.3. Nasality by stress

Figure 5 shows the proportion of vowels that are faithful in nasality to the Portuguese source form according to whether they are stressed (in Maxakalí) or not stressed (either in Maxakalí or in the Portuguese source form). Portuguese oral vowels behave similarly whether or not they are stressed ($p = .37$, non-significant), but for nasal vowels, we see a large difference in the likelihood that nasality will be maintained depending on stress ($p < .001$). Portuguese nasal vowels that end up in the stressed syllable in Maxakalí always retain their nasality in our sample, whereas if they are not stressed, they often lose their nasality. Note that because stress always occurs on the final syllable in Maxakalí, it is difficult to know for sure whether this effect is due to stress or due to word position. It is also worth noting that although the location of stress in Portuguese can vary, it is common for the stressed syllable in Portuguese to become the stressed

syllable in Maxakalí due to truncation, e.g. Mxk [nũˈbɛ̃β̃] ‘November’ < BP [noˈvẽbɾʊ]) (see Section 1). There are cases where there is a mismatch between the stressed syllable in Portuguese and the stressed syllable in Maxakalí (e.g. Mxk [mɛ̃ˈʤa] ‘table’ < BP [ˈmezə]), but such cases were rare in our sample. Thus, it is difficult to be certain whether the effect is due to the vowel being stressed in Maxakalí or due to it being stressed in the Portuguese source word (or both).

[INSERT FIGURE 5 HERE]

3.4. Nasality by place and manner of articulation

Table 1 shows the proportion of onsets that remain faithful in nasality to the Portuguese source form, according to their place and manner of articulation. Dorsal nasals are not allowed as onsets in either Maxakalí or Portuguese, so dorsals are always oral in this position.[11] In terms of place of

articulation, labial consonants are the most likely to have a change in nasality, particularly from oral to nasal. Coronals and palatals are less likely to change their nasality compared to labials.

[INSERT TABLE 1 HERE]

3.5. Summary of patterns in the corpus

To summarize, we find the following broad patterns in the Maxakalí loanword corpus collected:

1. Domain-based nasal agreement: There is absolute agreement in nasality within a rhyme. Nasality agreement *within* the syllable is stronger than across syllables.
2. There is a broad preference to preserve nasality, particularly in vowels and particularly in the stressed syllable.
3. There is a general preference to have a nasal vowel in the first syllable, especially if there is no onset.
4. Nasal vowels are avoided between two voiceless consonants.
5. Word-initial consonants always preserve the nasality or orality of their correspondent in Portuguese.
6. Labial consonants are more likely to be nasalized than other consonants.

Given that multiple (sometimes opposing) pressures appear to influence loanword adaptation in Maxakalí, these data call for computational modeling within a formal probabilistic framework. This approach provides us with a more explicit way to determine how much evidence is sufficient for a given constraint to be considered as holding in the adaptation process. To answer this, one needs statistical evaluation of the work done by proposed constraints, some of which combine

additively (see Halle & Keyser 1971: 176 for an early proposal) to determine the degree of wellformedness of a given loanword adaptation (relative to the alternatives). MaxEnt modeling provides a way of diagnosing, for instance, when a particular grammatical pressure is better conceived as a combination of two constraints, and it provides a way to decide between two overlapping constraints that can partially account for the same data, where one constraint might indeed render the other unnecessary. Finally, it allows one to figure out when a given analysis for loanwords undergenerates or overgenerates (used by Hayes, Wilson & Shisko 2012 for constraint sets proposed for metrical wellformedness of Shakespeare & Milton). We now turn to our analysis of the data in a MaxEnt framework.

4. MaxEnt analysis of Nasal Harmonization

In this section, we first provide an overview of MaxEnt grammars, justifying their applicability to Maxakalí loanword adaptation, and the procedure for selecting constraints through statistical model comparison. We then give an overview of the constraints that were put into the initial model. Lastly, we conclude with the set of constraints that were maintained in the final model and what they reveal about the relative strength of competing generalizations.

4.1. MaxEnt grammars

A MaxEnt grammar is a type of Harmonic Grammar (Legendre et al. 1990, Smolensky and Legendre 2006). Each constraint in the grammar is associated with a numerical weight. Output candidates are evaluated by taking the sum of the weighted violations that they incur across all

constraints. This evaluation method means that MaxEnt grammars, like other varieties of Harmonic Grammar, have the 'ganging up' property, whereby the combined violations of one or more weaker constraints can overpower a stronger constraint (Hayes and Wilson 2008, Pater 2009). Notably, MaxEnt grammars are probabilistic; after the weighted violations are summed, rather than selecting a single candidate as the winner, the EVAL component generates a probability distribution over all possible candidates for a given input. The probability of an output form is calculated by raising *e* (the base of the natural logarithm) to the negative summed weighted violations for the candidate, and then dividing by the total across all candidates for that input. This is what allows the model to capture variability in the output (e.g. Mxk [ijˈpuə̯] ~ [ĩj̃ puə̯] 'spur' < BP [isˈpɔɾə]). For in-depth overviews of MaxEnt grammars, see Goldwater and Johnson 2003, Hayes and Wilson 2008, and White 2013, 2017.

In our case, the MaxEnt framework provides a way of modeling the competing pressures on loanword adaptation, and it allows us to determine whether each constraint is statistically warranted in the model by submitting individual constraints to statistical testing. As such, we can avoid overfitting by only including constraints that are statistically justified given the data (Hayes, Wilson & Shisko 2012). Of greatest interest, discussed below, is finding out, among a host of potentially relevant factors encoded as probabilistic constraints within the model of loanword adaptation, which constraints actually do the most work in terms of explanatory coverage.

4.2. Input forms and output candidates

We used the Portuguese source word as the input form to the grammar. However, given that our interest is in nasality, we simplified by abstracting over several aspects of the adaptation not

directly related to nasality, including insertions, deletions, and truncations. In cases where segments were inserted or deleted, we took the input form to already have one correspondent for every segment in the observed Maxakalí form. For example, for the Maxakalí form [pãnãˈmæ̃ɲ̃] 'Flamengo (soccer team)', from Portuguese [flaˈmẽgʊ], we used the input form /falamẽg/, which already takes into account the epenthetic vowel and the truncated final vowel. We considered prenasalized oral stops to be oral for the sake of modeling (e.g. in Mxk [ᵑgaˈhaβ̰] ~ [gaˈhaβ̰] 'bottle' < BP [gaˈhafə], we treated [ᵑg] as [g]).

For each input form, we created a set of output candidates by varying whether each *voiced* segment (consonant or vowel) was oral or nasal. Final voiceless segments were assumed to become oral or nasal (voiced) glides in Maxakalí; in all other positions, voiceless segments in the input were assumed to remain voiceless (and therefore oral) in the output.[12] Using the same example as above, the Maxakalí form [pãnãˈmæ̃ɲ̃] was represented as an output candidate as [TṼNṼˈNṼN], where T is a voiceless consonant, Ṽ is a nasal vowel, and N is a nasal (voiced) consonant. This form was just one of the many output candidates considered for the input /falamẽg/. With two options (oral or nasal) for each voiced segment, the number of output candidates considered for a given input was 2^n, where *n* is the number of voiced segments. Thus, for /falamẽg/, the model considered $2^6 = 64$ output candidates. Though we did not directly model changes in place of articulation or manner (other than nasality), the model maintained this information in both the input and output forms, such that it could be referenced by constraints (e.g. faithfulness constraints specific to certain places of articulation, see below).

4.3. Implementation

The observed output forms (Maxakalí forms) were transcribed manually from audio recordings by a native speaker of Portuguese (the first author), and appropriate input forms were manually created for each of the observed outputs. We then created a Python script, which when given the list of input/output pairs, automatically (a) generated the set of output candidates for each unique input, (b) assigned violations to each output candidate, referencing the input form where appropriate, and (c) counted how many times each output candidate was observed in the corpus. We then used the MaxEnt Grammar Tool to run the MaxEnt models.[13] The MaxEnt Grammar Tool includes a Gaussian smoothing prior (see Goldwater & Johnson 2003); we used the default values for the prior ($\mu = 0$ and $\sigma^2 = 100000$ for all constraints), which means that the prior had very little effect on the weights learned.

4.4. Model comparison and constraint selection

We wanted to limit the *a priori* assumptions that we made about which constraints should be used to account for the data patterns. Therefore, we took the approach of including a large variety of constraints in an initial grammar, then using statistical model comparison to remove constraints that did not improve the model's fit to the data. Following Hayes, Wilson and Shisko 2012, we used likelihood ratio tests for the model comparison. The likelihood ratio test compares two models, where one model has a subset of the constraints included in the second model, and it determines whether the additional constraints improve the model's fit by a sufficient amount to be statistically warranted.

To find the best grammar, we used backwards stepwise comparison (i.e. a 'top-down' approach, as per the term used by Hayes, Wilson and Shisko 2012). We started with an initial grammar that contained all of the constraints that we considered. We first removed all constraints that received a weight of 0 in this initial grammar; these constraints could not possibly result in an improvement to the model's fit.[14] With the remaining constraints, we then removed one constraint at a time and compared the simpler model to the full model using a likelihood ratio test to gauge the extent to which each constraint improved the model's fit. The constraint that resulted in the least improvement to the model was then removed, and a new set of likelihood ratio tests was used to evaluate each of the remaining constraints. We continued this process, removing one constraint at a time, until all remaining constraints significantly improved the model's fit at an alpha level of .05.[15] At this point, we added each of the previously excluded constraints back into the final model one at a time to confirm that none of these constraints significantly improved the final model. We also added the full group of excluded constraints to the model at once (except those assigned a weight of 0) to confirm that the final model was indeed a better model than the initial model that we started with.

Likelihood ratio tests can only compare two models in a subset relationship, so this method is unable to compare two models with distinct sets of constraints. For instance, it is possible that a single more general constraint can do most of the work of two different specific constraints, a fact that can be obscured when only one constraint is removed from the model at a time. In cases where we suspected this might be the case, we compared the two models using two other measures of model fit that take into account the number of parameters, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). If the alternative model turned out to have a better AIC and BIC, then we used that model instead, and then continued with the backwards stepwise comparison using likelihood ratio tests. This occurred once in reaching the

final model below. Note that in all cases where we relied on likelihood ratio tests to make decisions about reducing the model, the AIC and BIC were also in agreement with the likelihood ratio tests.

Taking a statistical approach to constraint selection has two benefits (see also Hayes, Wilson & Shisko 2012). First, it provides a way to determine whether there is sufficient evidence in the data for a given constraint to be included in the model. To reach significance in the likelihood ratio test, a constraint must cause a sufficiently large increase in the likelihood of the data. The constraints can accomplish this either (a) by accounting for a subtle effect that is relevant to a large portion of the data, or (b) by accounting for a robust pattern, even if it is relevant only to a small subset of the data. A second benefit is to help in diagnosing when multiple constraints can account for the same pattern and selecting which constraint(s) to keep. If two constraints can account for the same aspect of the data, the model comparison process will determine whether both constraints are needed to account for the data, or whether just one is warranted. Constraints that have more general applicability and ones with fewer exceptions will be preferred over constraints that are highly specific or have many exceptions, all else being equal.

4.5. Overview of constraints included in the initial grammar

We included a total of 50 constraints in the initial grammar, keeping in mind that it is just as interesting to see which constraints were kept as to see which constraints were *not* kept in the end. In deciding upon the constraints to include, we considered previous accounts of Maxakalí nasal harmony, our own investigation of the data in the corpus, and other factors that we thought might be relevant. We summarize the constraints below.

4.5.1. Markedness constraints

We included 14 markedness constraints in the initial grammar. The first set of constraints enforces local agreement in nasality, meaning that only adjacent segments need to be considered in evaluating the constraint. These are formalized as AGREE[F] constraints (Pulleyblank 1997, Lombardi 1999, Baković 2000), defined in (7).

(7) AGREE[F]: Assigns one violation for each pair of adjacent segments that have different values for the feature [F].

In addition to the general AGREE[nasal], we also include four specific versions of the constraint that apply only in certain contexts. These are listed in (8).

(8) *Constraints enforcing local agreement in nasality*

a. AGREE[nasal]
b. AGREE[nasal]$_{VCD}$ (applies to voiced sounds only)
c. AGREE[nasal]-SYLL (applies only within a syllable)
d. AGREE[nasal]$_{VCD}$-SYLL (applies only to voiced sounds within a syllable)
e. AGREE[nasal]-RIME (applies only within a rhyme)

AGREE[nasal]$_{VCD}$ is a specific version of the constraint that only considers voiced segments; thus, the first two segments of the Maxakalí form [pũˈnɑ̃j̃] ‘FUNAI (National Agency for Indigenous Affairs)’ (< BP [fuˈnaɪ̯]) violate AGREE[nasal] but not AGREE[nasal]$_{VCD}$. (8c-e) are

versions of the constraint restricted to specific prosodic domains. AGREE[nasal]-SYLL applies to adjacent segments that are within the same syllable. Given the syllable structure of Maxakalí, this constraint is violated when there is a mismatch in nasality between a nucleus and its onset or a nucleus and its coda, but not when there is a mismatch, for instance, between a nucleus and the onset of the next syllable. For instance, the form [ka.mã.ˈdoɯ̨] 'horse' (< BP [kaˈvalʊ]) contains two violations of AGREE[nasal], but zero violations of AGREE[nasal]-SYLL because the two mismatches in nasality span syllable boundaries ([a.m] and [ã.d]). AGREE[nasal]$_{VCD}$-SYLL is a version that only evaluates adjacent voiced segments that are within the same syllable (i.e. it combines the restrictions in (8b) and (8c)).[16] Finally, AGREE[nasal]-RIME is restricted to the rhyme constituent, and thus is only violated when there is a mismatch between a nucleus and its coda.[17] One may note that a violation of AGREE[nasal]-RIME will always entail a violation of AGREE[nasal]-SYLL, which will itself always entail a violation of AGREE[nasal]. With this nested constraint set, it is guaranteed that a violation of AGREE[nasal] within a given domain will be penalized at least as much as a violation across the domain boundary (e.g. a violation within a syllable will be penalized at least as much as a violation across syllables).

We included three constraints motivating nasal harmony between syllables, shown in (9). These constraints assume an autosegmental tier (or some comparable representation) in which only the vowels are projected (Goldsmith 1979). The first constraint is a version of AGREE[nasal] that is evaluated only on the vowel tier. The other two are directional alignment constraints (Kirchner 1993, Archangeli & Pulleyblank 2002).

(9) *Constraints motivating nasal harmony across syllables*

a. AGREE[nasal]$_{VTIER}$ (restricted to vowel tier)

b. ALIGN[nasal]-L Assigns one violation for each vowel separating a feature [nasal] from the left edge of the word.

c. ALIGN[nasal]-R Assigns one violation for each vowel separating a feature [nasal] from the right edge of the word.

AGREE and ALIGN constraints have distinct properties when it comes to harmony (e.g. see Archangeli & Pulleyblank 2002, Finley 2008). AGREE constraints are symmetrical (motivating bidirectional spreading). They also have the 'sour grapes' property, which means that harmony will only occur if it is complete because partial spreading does not provide any improvement. On the other hand, ALIGN constraints are directional (motivating spreading in a single direction) and gradient, which means that the closer nasality can spread to the left (or right) edge of the word, the greater the improvement. We included both types of constraints so that the model could retain whichever provided the best fit to the data.

The final set of markedness constraints penalizes having nasality or orality in specific positions, motivated by our observations from the corpus data given in Section 3. We included one set of constraints that motivate having nasality within prominent positions, given in (10). We discuss the (partially language-specific) motivation for such constraints in Section 5.2.

(10) *Constraints promoting nasality in prominent positions*

a. *V_{ORAL}-σ_1 (penalizes oral vowels in the first syllable)

b. *#V_{ORAL} (penalizes word-initial oral vowels)

c. *#V_{ORAL}-ˈσ (penalizes oral vowels in the Mxk stressed syllable)

We also included a set of constraints that penalizes nasal vowels next to voiceless consonants, shown in (11). These constraints penalize a subset of the cases that the general AGREE constraints (the ones not limited to voiced segments) penalize. These were included because the data suggested a tendency for vowels to lose nasality (which is otherwise rare) specifically when surrounded by voiceless consonants. As mentioned above, constraints are additive in Harmonic Grammars, and the model selection process allows us to determine which of the partially overlapping constraints best account for the data, either individually or additively.

(11) *Constraints penalizing nasal vowels adjacent to voiceless consonants*

a. *TṼ (penalizes nasal vowels preceded by voiceless consonants)

b. *#TṼ (penalizes nasal vowels preceded by voiceless consonants at the beginning of the word)

c. *TṼT (penalizes nasal vowels between two voiceless consonants)

Summarizing, the markedness constraints included ones that motivated nasal agreement and ones that promoted or penalized nasality in particular positions. We now turn to the faithfulness constraints introduced into the model.

4.5.2. Faithfulness constraints

We included a total of 36 faithfulness constraints (Prince & Smolensky 2004[1993], McCarthy & Prince 1995) in the initial grammar. Given that we take the Portuguese form as the input form, the relevant correspondence relationship being evaluated by these faithfulness constraints is between the source word in Portuguese (as interpreted by Maxakalí speakers) and the borrowed surface form in Maxakalí (see Smith 2009). All of the faithfulness constraints fall into two general classes, IDENT and MAX. IDENT(nasal) is violated whenever a segment has a different value of [nasal] in the Maxakalí surface form than in the input form (i.e. in Portuguese). MAX(nasal) is violated only by a loss of nasality from the input form to the output form.[18] The use of MAX, in addition to IDENT, is motivated by the observation that Maxakalí speakers tend to maintain nasality already present in Portuguese when borrowing words, particularly in certain positions.

In addition to general IDENT(nasal) and MAX(nasal), we included versions of each constraint that apply specifically to consonants or vowels. These are shown in (12).

(12) *General Faithfulness constraints*

a. IDENT(nasal)

b. MAX(nasal)

---

c. IDENT(nasal)-C

(applies only to consonants)

d. MAX(nasal)-C

---

e. IDENT(nasal)-V

(applies only to vowels)

f. MAX(nasal)-V

We also included several positional faithfulness constraints (e.g. Beckman 1997, 1998, Casali 1997, Becker et al. 2012), which enforce the correspondence in nasality between input and output forms in specific prominent positions. We added constraints for all positions where we suspected there might be greater faithfulness, including onsets, word-initial position, stressed syllables (in both Maxakalí and Portuguese), and combinations of these positions (e.g. stressed onsets). The full list of positional faithfulness constraints included in the initial grammar is given in (13). Note that Maxakalí stress always occurs on the final syllable, so a constraint enforcing faithfulness within the Maxakalí stressed syllable is equivalent to one enforcing faithfulness within the final syllable.

(13) *Positional faithfulness constraints*

a. IDENT(nasal)-ONSET

(applies only to onsets)

b. MAX(nasal)-ONSET

---

e. IDENT(nasal)-WORDINITIAL

(applies only to word-initial position)

f. MAX(nasal)-WORDINITIAL

---

g.	IDENT(nasal)-STRESSEDSYLL$_{MXK}$	(applies only within the Maxakalí
h.	MAX(nasal)-STRESSEDSYLL$_{MXK}$	stressed syllable)
i.	IDENT(nasal)-STRESSEDV$_{MXK}$	(applies only to the vowel of the
j.	MAX(nasal)-STRESSEDV$_{MXK}$	Maxakalí stressed syllable)
k.	IDENT(nasal)-STRESSEDONSET$_{MXK}$	(applies only to the onset of the
l.	MAX(nasal)-STRESSEDONSET$_{MXK}$	Maxakalí stressed syllable)
m.	IDENT(nasal)-STRESSEDSYLL$_{SOURCE}$	(applies only within the Portuguese
n.	MAX(nasal)-STRESSEDSYLL$_{SOURCE}$	stressed syllable)
o.	IDENT(nasal)-STRESSEDV$_{SOURCE}$	(applies only to the vowel of the
p.	MAX(nasal)-STRESSEDV$_{SOURCE}$	Portuguese stressed syllable)
q.	IDENT(nasal)-STRESSEDONSET$_{SOURCE}$	(applies only to the onset of the
r.	MAX(nasal)-STRESSEDONSET$_{SOURCE}$	Portuguese stressed syllable)

Finally, we included faithfulness constraints that apply specifically to segments with certain places of articulation (14) or manners of articulation (15). Note that MAX is not relevant in the case of most manners of articulation, as it would only apply if the input segment were already nasal. Because MAX(nasal)-C already accounts for cases that would be covered by IDENT(nasal)-NASAL, the latter is not included.

(14) *Faithfulness constraints for specific places of articulation*

a.	IDENT(nasal)-LABIAL	(applies only to labials)
b.	MAX(nasal)-LABIAL	
c.	IDENT(nasal)-CORONAL	(applies only to non-palatal coronals)
d.	MAX(nasal)-CORONAL	
e.	IDENT(nasal)-PALATAL	(applies only to palatals)
f.	MAX(nasal)-PALATAL	
g.	IDENT(nasal)-DORSAL	(applies only to dorsals)
h.	MAX(nasal)-DORSAL	

(15) *Faithfulness constraints for specific manners of articulation*

a.	IDENT(nasal)-PLOSIVE	(applies only to input plosives)
b.	IDENT(nasal)-FRICATIVE	(applies only to input fricatives)
c.	IDENT(nasal)-LIQUID	(applies only to input liquids)
d.	IDENT(nasal)-RHOTIC	(applies only to input rhotics)
e.	IDENT(nasal)-GLIDE	(applies only to input glides)
f.	MAX(nasal)-GLIDE	(applies only to input glides)

4.6. Final grammar

Through model comparison, we saw a reduction from 50 constraints in the initial grammar to 19 constraints in the final grammar, including 7 markedness constraints and 12 faithfulness

constraints. The constraints in the final grammar are provided in (16). The p-value reported for each constraint comes from a likelihood ratio test comparing the final model to a simpler model in which the constraint in question was removed. In other words, it reflects whether including the constraint in the final model provides a significant improvement to the model's fit to the data. The weight assigned to each constraint generally corresponds to how exceptionless the constraint is in the corpus: higher weights are assigned to constraints with few exceptions, lower weights to constraints with many exceptions. Recall that a constraint with even a fairly low weight can reach significance in the model if the constraint has wide applicability (i.e. by being relevant to many forms in the corpus), whereas a constraint with a high weight might not reach significance if it explains only a couple of forms or if other more general constraints can explain the same data.

(16) *Constraints in final grammar*

MARKEDNESS CONSTRAINTS

Constraint		Weight	p-value
a.	AGREE[nasal]$_{VCD}$	1.47	< .001
b.	AGREE[nasal] –SYLL	1.68	< .001
c.	AGREE[nasal]-RIME	12.20	< .001
d.	ALIGN[nasal]-R	0.71	< .001
e.	*V_{ORAL}-σ_1	1.76	< .001
f.	*#V_{ORAL}	12.79	< .001
g.	*T$\tilde{V}$T	1.21	< .001

FAITHFULNESS CONSTRAINTS

Constraint		Weight	p-value
h.	MAX(nasal)-V	1.54	< .001
i.	IDENT(nasal)-WORDINITIAL	12.41	< .001
j.	MAX(nasal)-STRESSEDSYLL$_{MXK}$	2.27	< .001
k.	MAX(nasal)-STRESSEDV$_{MXK}$	12.16	< .001
l.	IDENT(nasal)-CORONAL	2.03	< .001
m.	IDENT(nasal)-PALATAL	2.74	< .001
n.	IDENT(nasal)-DORSAL	13.20	< .001
o.	IDENT(nasal)-PLOSIVE	1.03	< .001
p.	IDENT(nasal)-FRICATIVE	0.81	< .001
q.	IDENT(nasal)-RHOTIC	1.09	< .001
r.	IDENT(nasal)-GLIDE	2.84	< .001
s.	MAX(nasal)-GLIDE	10.17	.02

We turn to a discussion of which constraints were retained in the final model, and where relevant, which ones were not. We also discuss what the analysis means for our understanding of Maxakalí phonology.

5. Discussion

5.1. Nasal agreement

The domain-based hierarchy for nasal agreement that we saw above (Section 3.1) is reflected in the three local agreement constraints, (16a-c). AGREE[nasal]-RIME receives a large weight (12.20), reflecting the fact that nasal agreement is absolute within the rhyme. Outside of the rhyme, nasal agreement is not absolute, so AGREE[nasal]$_{VCD}$ and AGREE[nasal]-SYLL receive more modest weights (1.47 and 1.68, respectively). Recall that these weights are additive, so the sequence [ma] would violate both AGREE[nasal]$_{VCD}$ and AGREE[nasal]-SYLL (for a total of 3.15) whereas the sequence [a.mã] would violate only AGREE[nasal]$_{VCD}$ (for a total of 1.47) since the agreement violation does not involve two segments within a syllable.

The general AGREE[nasal] constraint was restricted to voiced sounds whereas the version applying within a syllable was not. This difference suggests that voiceless sounds are more likely to motivate the presence of an oral vowel within a syllable compared to across syllables. Of the markedness constraints specifically banning nasal vowels next to voiceless consonants, only *TṼT was retained in the final grammar, consistent with our observation (Section 3.2) that nasal vowels are avoided in Maxakalí when flanked on both sides by a voiceless consonant. Indeed, despite the general preference to maintain nasality on vowels in Maxakalí, we see that nasal vowels are often denasalized when a voiceless consonant occurs on both sides (e.g. Mxk [kuputaˈdo] ‘computer’ < BP [kõputaˈdoh]). Phonetically, the constraint could be motivated by the articulatory difficulty of TṼT sequences; they require lowering and rapidly raising the velum, while also maintaining a precise timing for the beginning and end of voicing (for a similar account of the *NC̨ constraint, see Huffman 1993, Pater 2004).

Constraints motivating nasal harmony between syllables played little role in the final grammar. In fact, ALIGN[nasal]-L and AGREE[nasal]$_{VTIER}$ received 0 weight in the initial model, indicating that they play no role in explaining the data. The 0 weight for ALIGN[nasal]-L is particularly noteworthy given that it represents the intuition expressed in work such as Rodrigues

(1981), which assumes leftward spreading of nasality across the word. ALIGN[nasal]-R was retained in the final grammar, but it was assigned a fairly low weight, suggesting that it plays a small role, at best, in determining the nasality of vowels.

Overall, we conclude that nasal agreement in Maxakalí is driven primarily by local agreement, with stronger agreement within smaller prosodic domains.

5.2. Initial syllables, faithfulness and a general preference for nasality

Word-initial consonants always retain the nasality/orality of their correspondent in the Portuguese source form. In our corpus, this is an absolute restriction, which is reflected in the high weight of IDENT(nasal)-WORDINITIAL (12.41). This restriction appears to be faithfulness in a particularly salient position, the initial onset of the word (Beckman 1997, 1998); presumably this position is especially salient in terms of making the borrowed form sound like the source form.

Beyond the initial consonant, there seems to be a general preference in Maxakalí for having nasality within the initial syllable. Vowels in word-initial position are very likely to be nasal in Maxakalí, whether or not they are nasal in the source form. In fact, we find many cases where there is spontaneous nasalization of initial vowels, meaning a word-initial oral vowel in Portuguese becomes nasalized in Maxakalí even without other sources of nasality in the word (e.g. Mxk [ã̲ˈgʊj] ‘August’ < BP [a̲ˈgostʊ]). As a result, *#V_{ORAL} receives a very high weight (12.79), which is necessary to offset the high weight of IDENT(nasal)-WORDINITIAL. In terms of explaining the motivation for this constraint, there are two extant possibilities. One, proposed by Rodrigues (1986) (and further developed in Sandalo & Abaurre 2010 for Pirahã and Fujimura 2010 for Kaingang) is that a lowered velum is a default articulatory setting in Maxakalí (and perhaps more broadly in lowlands South American indigenous languages), accounting for the

widespread phenomena of obstruent prenasalization as well as spontaneous nasalization. The second possibility is that onsetless vowels, particularly in the initial syllable, can achieve increased prominence by adding nasality (a strategy sporadically found for onsetless initial vowels in BP as well, e.g. [ĩˈgɾeʒə] ‘church’, cf. standard [iˈgɾeʒə]; Pinheiro 2019, though Pinheiro’s sample does not include such nasalization in any of the source words that undergo spontaneous left-edge nasalization in Maxakalí). Either explanation could motivate the strength of *#V_{ORAL}.

An exception to the pattern of nasalizing vowels in the first syllable is found when an initial oral vowel in Portuguese is followed by an oral coda; these vowels often remain oral in Maxakalí, despite being word-initial (e.g. Mxk [i̯i̯ˈpuə̯] ~ [ĩ̯ˈpuə̯] ‘spur’ < BP [i̯sˈpɔɾə]). Because nasal agreement in rhymes is absolute, nasalizing the vowel in such cases requires additionally nasalizing the coda glide, incurring a violation of IDENT(nasal)-GLIDE in addition to IDENT(nasal)-WORDINITIAL.

Vowels in the first syllable that are not word-initial (because there is an onset) are also more likely to be nasal compared to vowels in other positions, which is reflected by the moderate weight of *V_{ORAL}-σ_1 in the final model. This may provide an instance of domain generalization (Myers & Padgett 2014), in which a markedness constraint with strong phonetic motivation is generalized to a prosodic domain covering a wider range of instances than the original application, in this case of *#V_{ORAL} generalizing from word-initial vowels to all vowels in word-initial syllables.

As mentioned above (Section 3.2), the one case where we see an orality preference in the initial syllable is when the vowel is flanked on both sides by a voiceless consonant (e.g. BP cinquenta [sĩˈkʊ̃ẽtə] > Mxk [tɛikuˈɛ̃ə̃]; BP compadre [kũˈpadɾɪ] > Mxk [kuˈpɑə̯]; BP computador

[kõputaˈdoh] > Mxk [kuputaˈdo]; BP Santa Helena [sə̃teˈlenə] > Mxk [tɕataˈnɛ̃ɲ̃]). Consistent with this observation, *TṼT is retained in the final model.

5.3. Privileged status of nasality in Maxakalí

In general, we find that nasality tends to be preferentially preserved in Maxakalí, especially in vowels, whereas oral sounds frequently become nasal (see Section 3.2). This observation is borne out by the fact that MAX(nasal)-V is retained in the final model rather than a general IDENT constraint. Furthermore, we see that nasality is especially privileged in the Maxakalí stressed syllable, where nasal vowels are preserved without exception in our corpus (hence the high weight of 12.16 for MAX(nasal)-STRESSEDV_{MXK}).[19] The nasality of stressed nasal onsets also tends to be preserved, but to a lesser degree; hence MAX(nasal)-STRESSED$SYLL_{MXK}$ is retained with a modest weight in the final model.

As mentioned in Section 1, it is ambiguous based on the native Maxakalí vocabulary whether the nasality contrast is carried on the consonants or vowels in Maxakalí. Based on the loanword data, however, we can surmise that Maxakalí speakers traditionally interpreted the contrast as being on the vowel. This would explain why vowels (and especially stressed vowels) are more likely to maintain their nasality than consonants. Still, in cases where loanwords are realised with nasal disagreement, it provides limited evidence for a nasality contrast in both consonants and vowels. As more and more loanwords are borrowed into Maxakalí, we can imagine that a system in which nasality is fully contrastive in consonants and vowels might slowly develop (or might develop as a different stratum of vocabulary; e.g. Itô & Mester 1995, Smith 2018).

5.4. Place and manner effects

Several faithfulness constraints specific to certain places and manners of articulation are retained in the final model. These primarily mirror the patterns recorded in Table 1 above: places or manners that are more likely to change receive lower weights than those that are less likely to change. The constraint IDENT(nasal)-VELAR is assigned a very large weight due to the fact that [ŋ] is not allowed in onset position in Maxakalí, so any onset [g] from Portuguese is certain to remain oral. The constraint could be represented as a markedness constraint, *ŋ in onsets, with the same effect.

At least some of these place and manner constraints might be better conceived as constraints penalizing large changes as opposed to small changes (Steriade 2009; White 2014, 2017). For instance, [d] might be more likely to nasalize than the rhotic [ɾ] because it is more similar to [n], making it a less extreme change. Working out a precise account according to similarity is beyond the scope of this paper; here, we merely point it out as a possible alternative.

5.5. Effects across gender and age groups

The type of variability observed in loanword adaptation lends itself to investigating the potential role of demographic factors, such as age and gender, in conditioning the rates of repair strategies such as denasalization or harmony; indeed, earlier work on Maxakalí (Silva & Nevins 2014) found a partial role of age in conditioning the variable process of consonant lenition. We did not conduct a statistical analysis of nasalization and nasal harmony across the three demographic groups that formed part of the data collection here because there was an unequal representation of types across the various groups. However, for the sake of potential relevance to future research,

we wished to examine the rates of nasal agreement within each of the three domains (within the rhyme, within the syllable, across syllables) across the three age groups. We found that elders are more likely to have nasal agreement in borrowed words, both within syllables and across syllables (see Table 2). The rates of nasal agreement are comparable between the adult and youngster groups.

[INSERT TABLE 2 HERE]

These data suggest that that there has been a change in the willingness of speakers to tolerate nasal disagreement outside of the rhyme domain; specifically, younger speakers and adults are more likely to tolerate nasal disagreement than elder speakers. Put another way, younger speakers are more likely to remain faithful to the nasality in Portuguese, even if that results in a sequence that does not follow nasal agreement. The reason for this change could be greater exposure to Portuguese; though even the younger speakers of Maxakalí do not have bilingual fluency per se in Portuguese, they do have greater exposure to Portuguese through media and other sources, which may even reflect differences in attitude towards Portuguese lexical items. This increase in exposure may have led to an enhancement in the degree of faithfulness to the Portuguese source word for younger speakers. We found no reliable effects for gender.

Recall that as the data collection procedure was designed to elicit speakers' adaptation of particular Portuguese lexical items, not every speaker produced a loanword adaptation for every form (e.g. some produced fully native neologisms, such as *kõnãg ãxi-nã ax* (water cool-CAUS NMZ) for 'fridge', *pip xo-nix* (metal point-two) for 'scissors' and *ã-yĩy ax ã-xape pu* (2sg-talk NMZ 2sg-relative DAT) 'cellphone'). As a result, the data sets were not perfectly controlled for the

different age groups, and the comparison by age should be considered speculative. We leave it to future work to carry out a careful investigation of this potential change across generations.

5.6. Summary: Which Constraints Remained

Recall that through model comparison, we saw a reduction from 50 constraints in the initial model to 19 constraints kept in the final grammar, allowing one to see which version of a given constraint best modeled the patterns across the loanword corpus. Thus, the version of Agree(nasal) restricted to voiced segments favors nasalizing the vowel, whereas the version not limited to voiced segments does not favor nasalizing the vowel (as doing so would just create a new violation of Agree(nasal) within a TṼ sequence). In the data, nasalization is favored in such sequences, e.g. Port. [dokuˈmẽtʊ] > Mxk. [dukũˈmɛ̃ə̃], and these patterns constitute no violation of Agree(nasal)$_{VCD}$.

Of perhaps greatest interest is how the patterns of nasalization in loanword adaptation may also help disambiguate between a host of distinct but converging models of nasal harmony as applying in the native vocabulary, a conclusion that echoes work by Broselow et al 2012 and Smith 2018, among many others. Recall, for example, that while several models can account for nasalization harmony found in native vocabulary (by appeal to say ALIGN[nasal]-L), the MaxEnt model found a domain-based stepladder effect for loanwords, with greater agreement in smaller domains. While the model suggested a slight effect for ALIGN[nasal]-R, ALIGN[nasal]-L provided no benefit. In a similar manner, our modeling revealed that oral consonants may be preserved before nasal vowels (leading to the potential emergence of an oral/nasal contrast). In fact, this converges with limited evidence in the native phonology with the diminutive suffix, which

spreads nasality leftward, points in the same direction, e.g. [ʤok] ‘straight’, dim: [ʤõŋnãŋ], [dak] ‘dry’, dim.: [dãŋnãŋ]. The high-weighted constraint IDENT(nasal)-WORDINITIAL, regulating source-loan identity, may have convergent origins from the native phonology itself in regulating base-derivative identity in this case.

Finally, the result that *#V_{ORAL} induces spontaneous nasalization in loanwords may find converging evidence in the overall rarity of onsetless oral vowels within the native phonology itself, a trend that might go otherwise unnoticed. The preference for word-initial nasality both within loanwords and the native vocabulary (e.g. the frequent prenasalization of word-initial voiceless stops in Maxakalí; see Wetzels & Nevins 2018) may reinforce the overall conclusion in Rodrigues (1986) of a default articulatory position of a lowered velum as a language-specific setting. Interestingly, these spontaneously-nasalized word-initial vowels can then induce left-to-right harmony, counter to the trend previously posited for this language, but consistent with both initial position and stressed (final) position being privileged loci for nasality.

6. Overall Conclusions

At the outset of this paper, we examined whether nasal harmony applied differentially across the domains of rime, syllable, and the word in Maxakalí. Our findings were that indeed, harmony could be modeled as a local, segment-to-segment process, and one whose strength of application increased within smaller domains. In terms of directionality, we found no overall right-to-left bias in harmony, and if anything, a weak left-to-right tendency.

Maxakalí speakers have strategies for introducing and maintaining nasality in loanwords that depend on a host of factors, most importantly on vowels, in initial syllables, and in stressed syllables. These results add to an overall body of findings on the importance of prominent positions in maintaining contrasts (Beckman 1998, Smith 2002). Notably, other factors have

weight as well, and specific versions of such constraints (e.g. *TṼT vs *TṼ, or *V_{ORAL}-σ_1 vs *#V_{ORAL}) are arbitrated among, in terms of their degree of coverage and exceptions in an additive model. The modeling revealed no gain from a constraint driving word-level leftward harmony in loanwords, and indeed as there is no morphology or affixation involved in any of these loanwords, this language has the potential to reveal 'pure directionality' effects. By extension we may go far enough to conclude that no such constraint is needed in the native phonology; instead, nasal harmony is pairwise-adjacent, with syllable-constituent internal versions that are more strongly entrenched.

References

Archangeli, Diana, and Douglas Pulleyblank. 2002. Kinande vowel harmony: Domains, grounded conditions and one-sided alignment. *Phonology* 19.2:139-188.

Araújo, Gabriel Antunes. 2000. *Fonologia e morfologia da língua Maxakalí.* Masters' thesis, Universidade Estadual de Campinas.

Baković, Eric. 2000. *Harmony, dominance and control*. Ph.D. dissertation, Rutgers, The State University of New Jersey.

Becker, Michael, Andrew Nevins, and Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88:231–268.

Beckman, Jill. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14:1–46.

Beckman, Jill. 1998. *Positional faithfulness*. Ph.D. dissertation, University of Massachusetts Amherst.

Broselow, Ellen, Marie Huffman, Jiwon Hwang, Sophia Kao, and Yu-An Lu. (2012).

Emergent Rankings in Foreign Word Adaptations.Proceedings of the 30th West Coast Conference on Formal Linguistics, ed. Nathan Arnett and Ryan Bennett, 98-108.

Campos, Carlos Sandro (2009). Morfofonêmica e Morfossintaxe do Maxakalí. Doctoral dissertation, Univ. Federal Minas Gerais.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: which vowel goes? *Language* 73:493–533.

Cristófaro Silva, Thaïs. 2007. *Fonética e fonologia do português roteiro de estudos e guia de exercícios*. São Paulo: Ed. Contexto.

Finley, Sara. 2008. *The formal and cognitive restrictions on vowel harmony*. Ph.D. dissertation, Johns Hopkins University.

Fujimura, Osamu. (2010). Syllable structure of Kaingang: A C/D Model Interpretation. Presented at TCP, July 2010.

Goldsmith, John. 1979. *Autosegmental phonology*. New York: Garland.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Workshop on Variation within Optimality Theory, Stockholm University, 2003*.

Gudschinsky, Sarah, Popovich, H. & Popovich, F. (1970). Native Reaction and Phonetic Similarity in Maxakalí Phonology. *Language* 46: 77-88.

Gussenhoven, Carlos & Haike Jacobs. 2011. *Understanding Phonology*. London: Hodder Education.

Halle, Morris, and S. Jay Keyser. 1971. *English stress: Its form, its growth, and its role in verse*. Harper and Row, New York.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Hayes, Bruce, Colin Wilson, and Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88:691–731.

Huffman, Marie K. 1993. Phonetic patterns and nasalization and implication for feature specification. In Marie K. Huffman and Rena A. Krakow (eds.), *Phonetics and Phonology 5: Nasals, Nasalization, and the Velum*, 303–327. San Diego: Academic Press.

Itô, Junko & Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In Jill Beckman et al. (eds.), *Papers in Optimality Theory*, 181-209. Amherst: GLSA.

Kang, Yoonjung. (2011). Loanword phonology. In *Companion to Phonology*, eds. Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice, 2258-2282. Malden, MA.: Wiley Blackwell.

Kenstowicz, Michael 2003. The role of perception in loanword phonology: A review of Les emprunts linguistiques d'origine européenne en Fon. *Studies in African Linguistics* 32. 95 – 112.

Kirchner, Robert. 1993. *Turkish Vowel Harmony and Disharmony: An Optimality Theoretic Account*. Paper presented at the Rutgers Optimality Workshop I.

LaCharité, Darlene, and Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36.2: 223-258.

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Report CU-CS-465-90*. Computer Science Department, University of Colorado at Boulder.

Lombardi, Linda. 1998. *Evidence for MAX feature constraints from Japanese*. Ms. Rutgers Optimality Archive 247.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17:267–302.

Mateus, Mira Helena and D'Andrade, Ernesto. *The Phonology of Portuguese*. Oxford: Oxford University Press, 2000.

McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *Papers in optimality theory (University of Massachusetts occasional papers in linguistics 18),* 249–384. Amherst: University of Massachusetts Department ofLinguistics.

Myers, Scott & Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31: 399-433.

Nevins, Andrew. 2015. Review of Operstein (2010). *Phonology* 32: 307-317.

Operstein, Natalie. 2010. *Consonant structure and prevocalization*. Amsterdam: John Benjamins.

Pater, Joe. 2004. Austronesian nasal substitution and other NÇ effects. In John J. McCarthy (ed.), *Optimality Theory in phonology: A reader*, 271–289. Oxford and Malden, MA: Blackwell.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33:999– 1035.

Peperkamp, Sharon, and Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 367-370.

Piggott, G. & H. van der Hulst. (1997). Locality and the nature of nasal harmony. *Lingua* 103, 85-112.

Pinheiro, Paula. 2019. A Nasalização Espontânea no Portugues do Brasil: Uma Análise Multifatorial. Masters' thesis, Universidade Federal do Rio de Janeiro.

Prince, Alan, and Paul Smolensky. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Pulleyblank, Douglas (1997). Optimality Theory and features. In Diana Archangeli & Terry Langendoen (eds.), *Optimality Theory: an overview,* 59–101. Oxford: Blackwell.

Ribeiro, Eduardo. (2012). Final consonants in Maxakalí and their comparative status. *Liames* 12: 189-193.

Rodrigues, Aryon. 1981. Nasalização e fronteira de palavra em Maxakalí. *Anais do V Encontro Nacional de Linguistica.* Rio de Janeiro: Pontifícia Universidade Católica, pp 305-311.

Rodrigues, Aryon. 1986. Silêncio, pausa, e nasalização. *Anais do VIII Encontro Nacional de Linguistica.* Rio de Janeiro: Pontifícia Universidade Católica, pp 153-158.

Sandalo, F. & Abaurre, M. B. (2010). Orality spreading in Pirahã. *Liames* 10, 7–19.

Silva, Mário Coelho and Andrew Nevins. (2014). Ethnosociophonology and consonant lenition in Maxakalí. *LinguiStica* 10: 47-60.

Silva, Mário André Coelho. 2015. *A coda consonantal em Maxakalí*. Masters' thesis, Universidade Estadual de Campinas.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.

Singerman, Adam. (2016). Nasal Harmony and Phonotactic Well-formedness in Tupari. *International Journal of American Linguistics*. 82.4: 453-485.

Sistema de Informação da Atenção à Saúde Indígena (SIASI). *Dados populacionais indígenas por diversos parâmetros de análise*. Available at:

<http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/70-sesai/9518-siasi>. Accessed in September 11th, 2017.

Smith, Jennifer. 2002. Phonological augmentation in prominent positions. Doctoral disseration, UMass Amherst.

Smith, Jennifer. 2009. Source similarity in loanword adaptation: Correspondence Theory and the posited source-language representation. In Steve Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation,* 155-177. London: Equinox.

Smith, Jennifer. 2018. Impossible nativization effects and productivity in loanword phonology. Paper presented at OCP 2018, London, 14 January 2018.

Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG. Ms., Johns Hopkins University. http://roa.rutgers.edu/article/view/86.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind*. Cambridge: MIT Press.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson and Sharon Inkelas (eds.), *The nature of the word: Studies in honor of Paul Kiparsky*, 151–180. Cambridge, MA: MIT Press.

Thomas, Guillaume. (2014). A Split Analysis of Nasal Harmony in Mbyá. *LinguiStica* 10: 75-105.

Walker, Rachel. (2003). Reinterpreting transparency in nasal harmony. In Jeroen van de Weijer, Vincent J. van Heuven and Harry van der Hulst (eds.) *The Phonological Spectrum: Volume I: Segmental structure,* John Benjamins Publishing pp.37-72.

Wetzels, W. Leo. (2009). Nasal Harmony and the Representation of Nasality in Maxakali: Evidence from Portuguese Loans. In: Calabrese, A.; Wetzels, W. L. (Eds.). *Loan Phonology*. John Benjamins Publishing, p. 241-270.

Wetzels. W. Leo & Andrew Nevins. (2018). Prenasalized Stops vs. Postoralized

Stops: The Diverse Functions of Enhancement. *Language* 94.4: 834-866.

White, James. 2013. *Bias in phonological learning: Evidence from saltation*. Ph.D. dissertation, UCLA.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130:96–115.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93:1–36.

[INSERT APPENDIX HERE]

Figure 1. Proportion of adjacent Maxakalí consonant-vowel pairs that agree in nasality by domain, according to whether they agree (left panel) or disagree (right panel) in nasality in the Portuguese source form. Numbers on the bars show the total number of tokens in each group. (VC = vowel and coda within a rhyme; CV = onset and vowel within a syllable; V.C = vowel and onset across a syllable boundary)

Figure 2. Proportion of vowels and onsets that are faithful in nasality to the Portuguese source form, according to syllable position and the nasality of the sound in Portuguese. Numbers on the bars show the total number of tokens in each group.

Figure 3. Proportion of vowels in initial syllables that are faithful in nasality to the Portuguese source form, according to the type of onset and the vowel's nasality in Portuguese. Numbers on the bars show the total number of tokens in each group.

Figure 4. Proportion of vowels in initial syllables with voiceless onsets that are faithful in nasality to the Portuguese source form, according to the type of following consonant. Numbers on the bars show the total number of tokens in each group. (T = voiceless consonant.)

Figure 5. Proportion of vowels that are faithful in nasality to the Portuguese source form, according to stress. Numbers on the bars show the total number of tokens in each group.

[1*] We thank audiences at GLOW 41, Univ. Autonoma de Barcelona, Univ. Sheffield, the *IJAL* editors, associate editor, and anonymous reviewers, and Michael Becker, Adam Singerman, and Leo Wetzels for comments. Special thanks are due to Joshua Ho, who assisted in coding the data and in implementing the MaxEnt model.

[2] The following abbreviations are used throughout this paper: Maxakalí (Mxk), Brazilian Portuguese (BP), Consonant (C), Vowel (V), Voiceless Consonant (T), Nasal Consonant (N), Nasal Vowel (Ṽ), Maximum Entropy (MaxEnt), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Feature [F], Syllable (SYLL), Voiced (VCD), Vowel Tier (VTIER), Left (L), Right (R), as well as the Optimality Theoretic Constraints IDENT, MAX, AGREE, ALIGN.

[3] Final consonants in Maxakalí can further undergo lenition to become glides; see Gudschinsky, Popovich & Popovich (1970), Operstein (2010), Silva & Nevins (2014), & Nevins (2015) for discussion.

[4] Though interestingly, older speakers of Japanese are reported to produce [ʃitebaŋku], with the vowel being 'sacrified' to avoid affrication, parallel to /mã/ in the adaptation of *martelo* mentioned in the text.

[5] Recall that the group's autonym is [tɪjcmɨ̃ˈʔɨ̃ɰ̃ə̃], and they call their language [ɨ̃ɲĩ̃ʃˈʔɑj].

[6] In the phonetic transcriptions found throughout this paper, we include narrow detail that reflects allophonic vowel quality beyond the five phonemic vowel qualities noted above, as well as final consonant lenition, but these factors are not relevant for the oral/nasal contrast that is our central research question.

[7] In our data, only the form with a medial nasal consonant, that is, [pɛ̃ˈɲõŋ], occurs for this word.

[8] The Fisher's exact tests reported throughout this section were conducted in R using the fisher.test() function. For each plot shown, we first conducted a test on the full table of data represented in the plot, obtaining a Monte Carlo simulated p-value (which were simulated due to the large size of the data tables). These overall tests were significant for all of the plots presented here (all $p < .001$); we do not report on the overall tests further. For the pairwise comparisons of interest that we report, we used a Bonferroni-corrected alpha level based on all reasonable pairwise comparisons in the data table. The p-values reported for pairwise comparisons were calculated exactly (rather than simulated).

[9] Tokens where rhyme segments disagreed in the Portuguese source word come from cases of truncation, in which an onset in Portuguese became the rhyme of the final syllable in Maxakalí (e.g. Mxk [nũˈbɛ̃β̃] 'November' from BP [noˈvẽbɾʊ]).

[10] The absolute faithfulness of initial-syllable nasal vowels not preceded by a voiceless onset should be taken tentatively, given the low number of tokens in those groups. Fisher's exact tests comparing nasal vowels preceded by a voiceless onset and nasal vowels preceded by other onset types are only borderline significant or trending towards significance at a Bonferroni-corrected alpha level of .003, again due to the small number of tokens (voiceless vs. no onset, $p = .003$; voiceless vs. voiced oral onsets, $p = .058$; voiceless vs. nasal onsets, $p = .006$).

[11] Due to phonotactic gaps, there are gaps in the cells of Table 1. Moreover, because we selected the loanwords from previous research, as mentioned in Section 2, and because not all speakers knew each of these loanwords, there are some cells with smaller numbers of tokens.

[12] This turns out to be a simplification, because final voiceless consonants sometimes remain voiceless rather than becoming glides in Maxakalí, but is a reasonable simplification given our sample.

[13] The MaxEnt Grammar Tool software was developed by Colin Wilson and Ben George, and made available for public use by Bruce Hayes at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

[14] Note that if multiple constraints can account for the same aspect of the data, the MaxEnt model will spread the weight out amongst these constraints, so a weight of 0 means that the constraint provides no benefit in terms of explaining the data.

[15] A reviewer inquired why we did not apply any type of correction for multiple comparisons to this alpha value. As this analysis does not consist of a traditional set of independent statistical tests, it is not clear that a correction is appropriate in this case. Even so, we note that all of the constraints retained in the final model except one have a p-value less than .001 (see Section 4.6), so even given the application of a very conservative Bonferroni correction for 50 constraints (resulting in a corrected alpha level of .05/50 = .001), all of the same constraints would be retained except one. Thus, there would be little difference in the overall conclusions.

[16] AGREE[nasal]$_{VCD}$-SYLL could be considered a conjoined version of constraints (8b) and (8c) (e.g. Smolensky 1995).

[17] We have not included a version of AGREE(nasal)-RIME restricted to voiced segments because voiceless segments generally do not occur in the coda position in our sample, with the occasional exception of [h], which we have abstracted over here.

[18] Using MAX(nasal) to penalize a loss of nasality assumes a representation where [nasal] is treated as an autosegment (Goldsmith 1979) or some other entity that can be deleted or retained independently from the full segment (e.g. Lombardi 1998). An IDENT constraint

asymmetrically penalizing a change from [+nasal] to [–nasal] would work equally well in this case.

[19] Due to truncation, it is most often the case that the stressed syllable in Portuguese remains the stressed syllable in Maxakalí. In our sample, there are only a handful of exceptions to this generalization. Given this overlap, using constraints referring to the Portuguese (source) stressed syllable rather than the Maxakalí stressed syllable results in almost as good a fit to the data, making it difficult to know with certainty whether it is the stress of the source form, the stress of the surface form, or some combination that is driving the stress effect.

Table 1. Proportion of onset consonants faithful in nasality according to place and manner of articulation.

	Labials	Coronals	Palatals	Dorsals
Plosives	0 (0/18)	.98 (84/86)	1 (4/4)	1 (20/20)
Fricatives	.53 (31/59)	--	.80 (73/91)	--
Nasals	.85 (82/98)	.90 (55/61)	1 (2/2)	--
Laterals	--	.78 (66/85)	--	--
Rhotics	--	.86 (64/74)	--	--
Glides (oral in BP)	--	--	--	--
Glides (nasal in BP)	--	--	1 (23/23)	--

Table 2. Rates of nasal agreement in each of the three domains, according to age group. (VC = within a rhyme; CV = onset and vowel within a syllable; V.C. = vowel and following onset in different syllables)

	VC	CV	V.C
elders	100%	81%	49%
adults	100%	73%	41%
youngsters	100%	76%	39%

Figure 1. Proportion of adjacent Maxakalí consonant-vowel pairs that agree in nasality by domain, according to whether they agree (left panel) or disagree (right panel) in nasality in the Portuguese source form. Numbers on the bars show the total number of tokens in each group. (VC = vowel and coda within a rhyme; CV = onset and vowel within a syllable; V.C = vowel and onset across a syllable boundary)

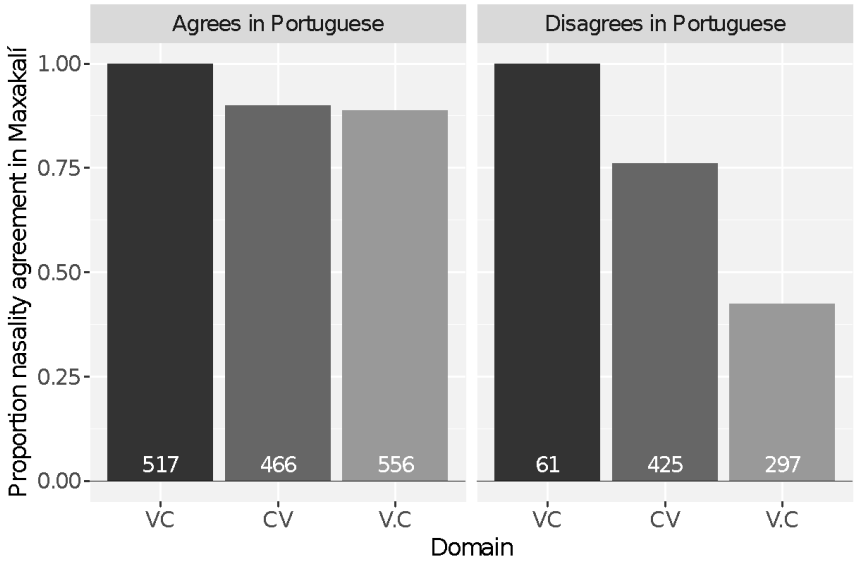


Figure 2. Proportion of vowels and onsets that are faithful in nasality to the Portuguese source form, according to syllable position and the nasality of the sound in Portuguese. Numbers on the bars show the total number of tokens in each group.

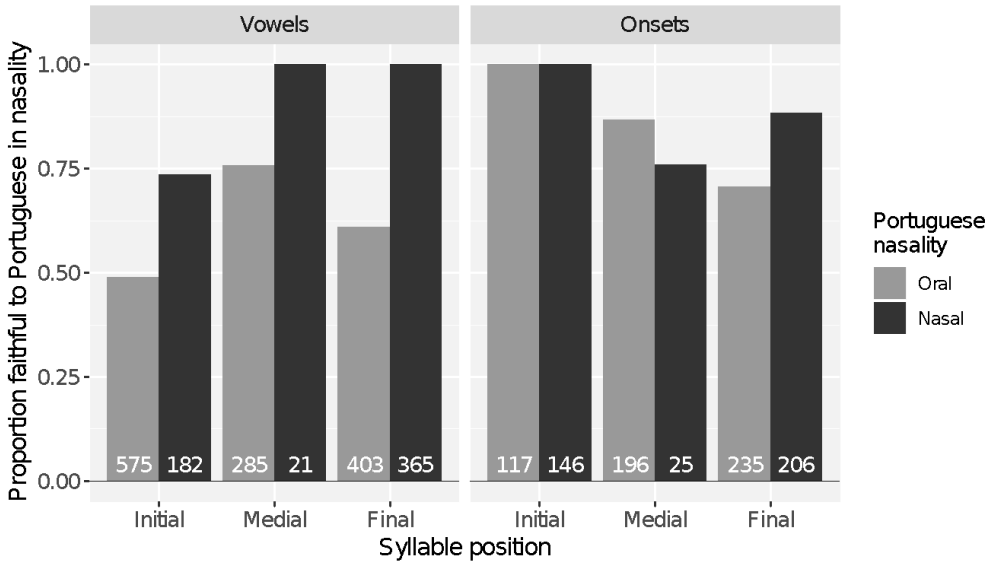


Figure 3. Proportion of vowels in initial syllables that are faithful in nasality to the Portuguese source form, according to the type of onset and the vowel's nasality in Portuguese. Numbers on the bars show the total number of tokens in each group.

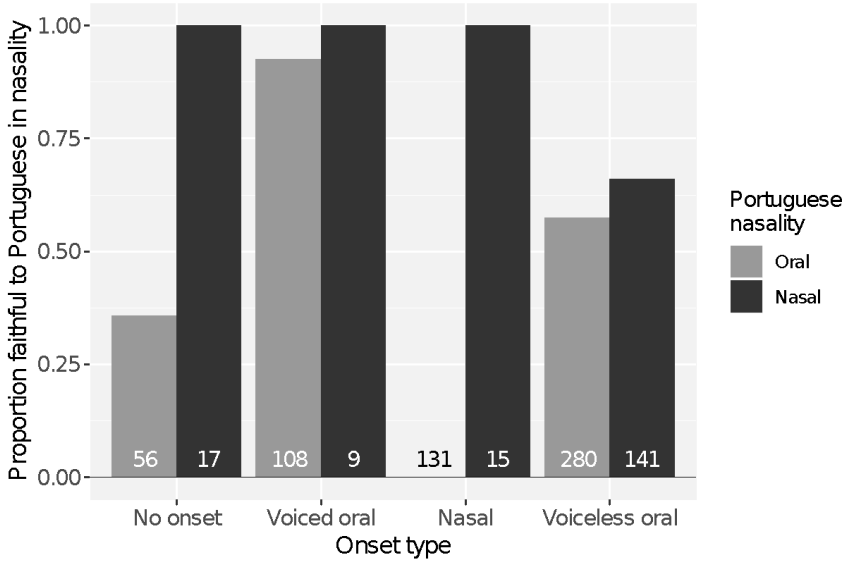


Figure 4. Proportion of vowels in initial syllables with voiceless onsets that are faithful in nasality to the Portuguese source form, according to the type of following consonant. Numbers on the bars show the total number of tokens in each group. (T = voiceless consonant.)

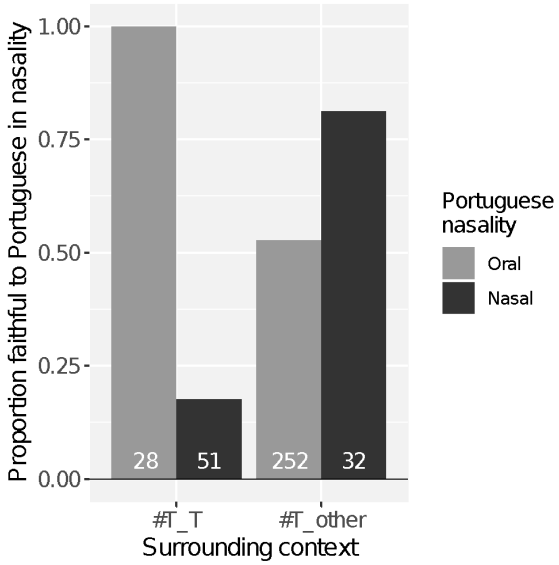
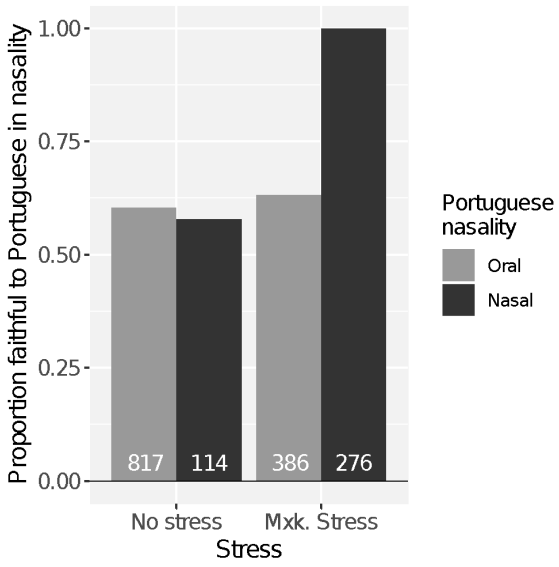


Figure 5. Proportion of vowels that are faithful in nasality to the Portuguese source form, according to stress. Numbers on the bars show the total number of tokens in each group.



Appendix of all loanword adaptations

Orthographic Portuguese	Portuguese Phonetic Transcription	Orthographic Maxakalí	Maxakalí Phonetic Transcription	Translation	Speakers for Each Variant (or Both)
remédio	[hɛˈmɛʤʊ]	hemẽn	[hɛˈmɛ̃n̪] ~ [hɛ̃ˈmɛ̃n̪]	medicine	16,1 (+ 1 for both)
Rio de Janeiro	[ˌhiʊ̯ʤiʒaˈnerʊ]	Hipyanẽn	[hiβ̪dzaˈnɛ̃n̪] ~ [hiβ̪dzãˈnɛ̃n̪]	Rio de Janeiro (city)	4,10
espelho	[isˈpeʎʊ]	ĩypex	[ɪjˈpæj] ~ [ĩ̞jˈpæj]	mirror	0,14 (+ 1 for both)
espora	[isˈpɔɾə]	ixpot	[ɪjˈpʊn̪] ~ [ĩ̞jˈpʊn̪]	spur	5,1
Estados Unidos	[isˌtaduzuˈnidʊs]	Ixtayunĩn	[ɪjtadzuˈnĩβ̞] ~ [ɪjtadzũˈnĩβ̞]	USA	3,1
cavalo	[kaˈvalʊ]	kãmãnok	[kamãˈdoɯ̞] ~ [kãmãˈdoɯ̞]	horse	4,9
caminhão	[kamĩˈɲãõ̞]	kãmĩyãm	[kamĩˈɲãβ̞] ~ [kãmĩˈɲãβ̞]	truck	7,7
computador	[kõputaˈdoh]	kopotano	[kuputaˈdo] ~ [kũputaˈdo]	computer	1,3
sabão	[saˈbãõ̞]	xãmããm	[tɕaˈmãβ̞] ~ [tɕãˈmãβ̞]	soap	2,16
cinquenta	[sĩˈkʊ̯ẽtə]	xikoẽn	[tɕikuˈɛ̃n̪] ~ [tɕĩkuˈɛ̃n̪]	fifty	1,1
canivete	[kãniˈvɛʧi]	kunimet	[kɨdiˈbɛn̪] ~ [kĩnĩˈbɛn̪]	pocket knife	6,4
Flamengo	[flaˈmẽgʊ]	Panamẽy	[padaˈmæ̃j̃] ~ [panãˈmæ̃j̃]	Flamengo (soccer team)	1,1
Santa Helena	[sə̃teˈlenə]	Xãtãnẽn	[tɕataˈnɛ̃n̪] ~ [tɕatãˈnɛ̃n̪] ~ [tɕãtãˈnɛ̃n̪]	Santa Helena (town)	8,6,3

janela	[ʒaˈnɛlə]	yanẽn	[dzaˈnɛ̃ñ̪] ~ [dzãˈnɛ̃ñ̪]	window	2,11
março	[ˈmahsʊ]	mãy	[ˈmɑ̃j̃]	March	
banheiro	[bə̃ˈj̃erʊ]	maet	[ˈbɐ̃ʔɛn̪]	toilet	
banco	[ˈbə̃kʊ]	mug	[ˈbɔ̃ŋ]	bank	
boi	[ˈboɪ̯]	mox	[ˈbʊj]	ox	
junho	[ˈʒũj̃ʊ]	yoy	[ˈdzʊ̃j̃]	June	
conto	[kõˈtʊ]	kõn	[ˈkõñ̪]	old Brazilian currency	
meia	[ˈmeɪ̯ə]	mẽ	[ˈmɐ̃j̃]	sock	
manga	[ˈmə̃gə]	mãg	[ˈmɐ̃ŋ]	mango	
Mário	[ˈmaɾɪ̯ʊ]	Mãn	[ˈmɑ̃ñ̪]	Mário (name)	
mãe	[ˈmə̃ɪ̯̃]	mãy	[ˈmɑ̃j̃]	mother	
maio	[ˈmaɪ̯ʊ]	mãy	[ˈmɑ̃j̃]	May	
nove	[ˈnɔvɪ]	nõm	[ˈnõβ̞̃]	nine	
pão	[ˈpə̃ʊ̯̃]	pãm	[ˈpɐ̃β̞̃]	bread	
cem	[ˈsẽɪ̯̃]	xẽy	[ˈtɕɐ̃j̃]	hundred	
Sandro	[ˈsə̃drʊ]	Xãn	[ˈtɕɑ̃ñ̪]	Sandro (name)	
Santos	[ˈsə̃tʊs]	Xãn	[ˈtɕɑ̃ñ̪]	Santos (city)	
cinco	[ˈsĩkʊ]	xĩy	[ˈtɕĩj̃]	five	
ônibus	[ˈõnɪbʊs]	õn	[ˈõñ̪]	bus	

agosto	[aˈgostʊ]	ãgox	[ãˈgʊj]	August	
açúcar	[aˈsukəh]	ãxok	[ãˈtɕoʊɥ]	sugar	
avião	[aviˈə̃ɰ̃]	amiyãm	[abĩˈɲãβ̞̃]	airplane	
Andrew	[ə̃ˈdɾeʊ̯]	Ãnenep	[ãdɛˈdɛβ̞̃]	Andrew	
armário	[aɦˈmarɻʊ]	ahmãn	[ahˈmɑ̃ɰ̃]	cupboard	
Belo Horizonte	[bɛloɾiˈzõʧi]	Menaniyon	[bɛdadiˈdʑũɰ̃]	Belo Horizonte (city)	
laranja	[laˈɾə̃ʒə]	nanuy	[daˈd̪ĩ̃ʃ]	orange	
Ladainha	[ladaˈĩ̃jə]	Nanaĩy	[dadaˈʔĩ̃ʃ]	Ladainha (town)	
dezembro	[deˈzẽbrʊ]	neyem	[dɛˈdʑɛ̃β̞̃]	December	
limão	[liˈmə̃ɰ̃]	nimam	[dibãβ̞̃]	lime	
documento	[dokuˈmẽtʊ]	nokomẽn	[dukũˈmɛ̃ɰ̃]	document	
janeiro	[ʒaˈneɾʊ]	yanẽn	[dʑaˈnɛ̃ɰ̃]	January	
dia de domingo	[ʤiədʒiduˈmĩgʊ]	yenomiy	[dzɛduˈbĩ̃ʃ]	day of Sunday	
garrafa	[gaˈhafə]	gahap	[gaˈhaβ̞̃]	bottle	
ganhou	[gə̃ˈj̃o]	gãyõ	[gã̠ˈɲõ]	to win	
governo	[guˈveɦnʊ]	gohet	[guˈhɛɰ̃]	government	
Rosângela	[hoˈzə̃ʒelə]	Hoyay	[huˈdʑɑ̃ʃ]	Rosângela (name)	
cangalha	[kə̃ˈgaʎə]	kãgax	[kã̠ˈgɑj]	yoke	
caldeirão	[kaʊ̯deˈɾə̃ɰ̃]	kanenam	[kadɛˈd̪ãβ̞̃]	cauldron	

carneiro	[kaɦˈnerʊ]	kahnẽn	[kahˈnɛ̃ɲ̃]	sheep	
calção	[kaʊ̯ˈsə̃ʊ̯̃]	katxãm	[kɑɐ̯̃ˈtɕãβ̞̃]	shorts	
quinhentos	[kĩˈj̃ẽtʊs]	kĩyẽn	[kĩˈɲɛ̃ɲ̃]	five hundred	
quinta-feira	[ˌkĩtəˈferə]	kĩtapet	[kĩtaˈpɛɐ̯̃]	Thursday	
comadre	[kuˈmadrɪ]	kõmãn	[kũˈmɑ̃ɲ̃]	kid's godmother	
comércio	[kɔˈmɛhsɪ̯ʊ]	kõmẽn	[kũˈmɛ̃ɲ̃]	shop (Mxk: city)	
compadre	[kũˈpadrɪ]	kopat	[kuˈpɑɐ̯̃]	kid's godfather	
colchão	[koʊ̯ˈʃə̃ʊ̯̃]	kotxãm	[kʊɐ̯̃ˈtɕãβ̞̃]	mattress	
Marina	[maˈɾĩnə]	Mãnĩn	[mãˈnĩj̃ɲ̃]	Marina (name)	
maçã	[maˈsə̃]	mãxã	[mãˈtɕã]	apple	
martelo	[mahˈtɛlʊ]	mãhtet	[mãˈtɛɐ̯̃]	hammer	
Maíza	[maˈizə]	Mãip	[mãˈʔiβ̞̃]	Maíza (name)	
macarrão	[makaˈhə̃ʊ̯̃]	mãkãhãm	[mãkãˈhãβ̞̃]	pasta	
maracujá	[maɾakuˈʒa]	mãnãkuya	[mãnãkuˈdʑa]	passion fruit	
Maxakalí	[maʃakaˈli]	Mãxakani	[mãtɕakaˈdi]	Maxakalí	
mesa	[ˈmezə]	mẽya	[mɛ̃ˈdʑa]	table	
Minas Gerais	[ˌmĩnəʒɛˈɾaɪ̯s]	Mĩnãyenax	[mĩnãdʑɛˈdɑj]	Minas Gerais (state)	
moça	[moˈsə]	moxa	[mõˈtɕa]	girl	
motor	[moˈtoh]	mõtok	[mũˈtoɯ̟̃]	engine	

motorista	[motoˈɾistə]	mõtonix	[mũtuˈdɪj]	driver	
na feira	[naˈferə]	nãpet	[nãˈpɛ̰]	at the fair (Mxk: 'fair')	
novembro	[noˈvẽbɾʊ]	nõmem	[nõˈbɛ̃β̞̃]	November	
no posto	[nuˈpostʊ]	nõpox	[nũˈpʊj]	at the (indigenous) post	
Oliveira	[oliˈverə]	Õnĩmet	[ũnĩˈbɛ̰]	Oliveira (surname)	
fazendeiro	[fazẽˈderʊ]	payenet	[padʑɛ̃ˈdɛ̰]	farmer	
peneira	[peˈnerə]	pẽnẽn	[pɛ̃ˈnɛ̃ɲ̃]	sieve	
feijão	[feˈʒə̃ʊ̃]	pẽyõg	[pɛ̃ˈɲõŋ]	beans	
presidente	[pɾeziˈdẽʧi]	peneyinen	[pɛdɛdʑiˈdɛ̃ɲ̃]	president	
fogão	[fuˈgə̃ʊ̃]	pogãm	[puˈgãβ̞̃]	stove	
Funai	[fuˈnaɪ̯]	Põnãy	[pũˈnɑ̃j̃]	Bureau of Indian Affairs	
'tá devendo	[tadeˈvẽnʊ]	tanemẽn	[tadɛˈmɛ̃ɲ̃]	to owe	
sandália	[sə̃ˈdalɪ̯ə]	xãnax	[tɕãˈdɑj]	sandals	
semana	[seˈmənə]	xẽmãn	[tɕɛ̃ˈmɑ̃ɲ̃]	week	
setembro	[seˈtẽbɾʊ]	xetẽm	[tɕɛˈtɛ̃β̞̃]	September	
segunda-feira	[siˌgũdəˈferə]	xigõnãpet	[tɕigũnãˈpɛ̰]	Monday	
São Paulo	[sə̃ʊ̯̃ˈpaʊ̯lʊ]	Xõpat	[tɕũˈpɑɲ̰]	São Paulo (city)	
Teófilo Otoni	[ʧɔflоˈtõnɪ]	Xopotõn	[tɕupuˈtũɲ̃]	Teófilo Otoni (city)	
telefone	[teleˈfõnɪ]	tenepõn	[tɛdɛˈpũɲ̃]	phone	

televisão	[televiˈzɐ̃ʊ̯̃]	tenemiyam	[tɛdɛbiˈdzɐ̃β̞̃]	TV	
trem-de-ferro	[ˌtɾɐ̃ɪ̯̃dʒiˈfɛhʊ]	tẽnẽynipe	[tɛ̃nɐ̃ɟ̃diˈpɛ]	train	
tomate	[tuˈmaʧi]	tõmãn	[tũˈmɑ̃ɐ̯̃]	tomato	
oitocentos	[oɪ̯tuˈsẽtʊs]	oxxẽn	[ʊjˈtɐɛ̃ɐ̯̃]	eight hundred	
oitenta	[oɪ̯ˈtẽtə]	oxtẽn	[ʊjˈtɛ̃ɐ̯̃]	eighty	