

ARTICLE

A corpus-based study of variation in and extension of two Paraguayan Guaraní nasalisation patterns

Katherine R. Russell 

Linguistics, University of California, Berkeley, USA.
Email: katherine.russell@berkeley.edu

Received: 1 August 2022; **Revised:** 30 June 2023; **Accepted:** 30 May 2024

Keywords: Paraguayan Guaraní; nasal harmony; variation; corpus phonology; suppletive allomorphy

Abstract

In Paraguayan Guaraní (PG), nasalisation processes affect material to both the left and right of a stressed nasal vowel. While some prior literature has claimed that bidirectional harmony is active in the language, others have noted that progressive nasalisation appears to be morpheme-specific and likely dependent on a different mechanism from regressive nasal harmony. Recent work shows that Spanish-origin lexical items participate in regressive nasal harmony, but the interactions of etymological origin and progressive nasalisation remain unclear. Drawing on a corpus of 26 sociolinguistic interviews as well as elicitation with native speakers of PG, I argue that the mechanisms underlying the two types of nasalisation in the language are in fact different. I propose that PG regressive nasalisation is best analysed as productive nasal harmony, while progressive nasalisation represents a case of morpheme-specific allomorphy. Additionally, though the PG pattern of regressive nasal harmony has been extended to items of Spanish origin, this is not the case for progressive nasalisation. This corpus study provides insight into the specific factors that condition variation in nasalisation processes, contributing to a growing literature investigating variable application of harmony.

Contents

1. Introduction	2
2. Background	3
2.1. PG morphophonology	3
3. Nasality and nasalisation in Paraguayan Guaraní	5
3.1. Nasality	5
3.2. Nasalisation	6
3.3. Interim summary	9
4. Spanish-origin items	10
4.1. Phonotactic adaptation	10
4.2. Nasality	11
4.3. Nasalisation	12
5. Quantifying variation	13
5.1. Methods	14
5.2. Findings	16
6. Discussion	21
7. Conclusion	24
A. Abbreviations	25
B. Figures	26
References	31



1. Introduction

Two directions of nasalisation – regressive (leftward) and progressive (rightward) – are attested in Paraguayan Guaraní (referred to henceforth by the initialism PG; ISO: gug), a Tupí–Guaraní language spoken by around 6.5 million people, primarily in Paraguay. Segments preceding a stressed nasal vowel are realized as anticipatorily nasalised: this is visible in (1a), where nasality spreads leftward from the root-final stressed vowel /ẽ/ in the root ‘arrive’. Nasalisation of a suffix or enclitic is also attested, triggered by the same root-final stressed vowel: the enclitic meaning ‘until’ surfaces with an initial nasal in (1a), as opposed to the initial voiceless stop of the same morpheme in an oral context like (1b).¹

- (1) a. mbaʔe've nd-o-j'ko=j *ã-ŋ^wã'hẽ=mẽve*
 nothing NEG-3.A-be=NEG ISG.A-arrive=until
 ‘Nothing happened until I arrived’. 20210301_mcg
- b. mbaʔe've nd-o-j'ko=j a-'ha=*peve*
 nothing NEG-3.A-be=NEG ISG.A-go=until
 ‘Nothing happened until I left’.

Drawing on data like that in (1), some previous work has claimed that PG represents a case of bidirectional harmony (Lunt 1973; Goldsmith 1976). Others have noted, however, that progressive nasalisation seems morpheme-specific (Lapierre & Michael 2018; Estigarribia 2021): I investigate these conflicting assessments here. This work also addresses the interactions of etymological origin and application of nasalisation. Spanish-origin roots are extremely frequent in PG and, due to differences in phonotactic constraints between Spanish and PG, provide an interesting lens through which to view PG nasalisation patterns. Regressive nasal harmony can be triggered by a nasal consonant within a Spanish-origin root (Russell 2022a); however, the question of whether Spanish-origin roots may trigger progressive nasalisation as well has not yet been addressed in the literature.

In this article, I tackle the question of whether regressive and progressive nasalisation processes in PG are in fact distinct, and if so, what different mechanisms underlie the two types of nasalisation. I examine the two directions of nasalisation and their interactions with root etymological origin through a corpus study of actual application rates, presenting the first quantitative study of nasalisation rate in PG. I find that direction does in fact affect application of nasalisation processes, and that nasalisation rate is influenced by different factors for each direction of nasalisation. Variation in the regressive nasalisation triggered by a root is influenced by factors including the morphosyntactic status of the affix and the token frequency of the root. Crucially, the process of regressive nasal harmony has been extended to Spanish-origin items, to the point that etymological origin of the root is not a significant predictor of application of nasal harmony. However, root etymological origin does greatly impact progressive nasalisation rate. I propose that regressive nasalisation is best analysed as (productive) nasal harmony, while progressive nasalisation represents a case of morpheme-specific phonologically conditioned suppletive allomorphy.

This article is structured as follows: in §2, I provide background on PG and its morphophonology. I then present an overview of nasality and nasalisation in PG in §3, including the two nasalisation processes in question: regressive nasal harmony and morpheme-specific progressive nasalisation. I introduce their interactions with Spanish-origin items in §4. Next, I present and briefly discuss the findings of a quantitative corpus study of nasalisation in PG in §5. I connect these findings to the

¹ All PG examples are represented using a three-line gloss method. First, the word or sentence is spelled in its phonetic representation in the International Phonetic Alphabet (IPA). Spanish-origin morphemes are italicised. These graphic representations are then followed by morphemic glossing (abbreviations used are listed in Appendix A) and English translations. Example sentences from primary data are noted with a reference consisting of the date in YYYYMMDD format and a three-letter code identifying the consultant. When multiple sentences within a single example come from the same session, I provide the code only for the first sentence. All materials collected are publicly available online through the California Language Archive (Gómez *et al.* 2020) at <https://doi.org/doi:10.7297/X2PR7TNF>.

larger picture of PG nasalisation and variation in vowel harmony cross-linguistically in §6, then briefly conclude in §7. Data presented here, unless otherwise cited, come from two sources: elicitation data collected remotely with two native speakers of PG (Gómez *et al.* 2020), and a corpus of sociolinguistic interviews conducted in Paraguay (Bittar Prieto 2021).

2. Background

PG is one of the most widely spoken Indigenous languages of the Americas: indeed, Paraguay is the only American nation where an Indigenous language has survived as a majority language spoken by the non-Indigenous population (Estigarribia 2020). Spanish colonisers arrived in what is now Paraguay in the early 1500s, but never had a large administrative presence and did not enforce the exclusive use of Spanish. The few Spanish men that lived in the region generally married ethnically Guaraní women, who raised their children speaking Guaraní (Morínigo 1959). The use of the Guaraní language has been a strong symbol of national identity for centuries, particularly as a way for Paraguayans to differentiate themselves from their neighbours in times of conflict (Estigarribia 2015). Today, PG is an official language of Paraguay, along with Spanish (*Constitución Política de la República de Paraguay*, article 140). It has been present in all schools in the country, either as the primary language of instruction or as a separate subject, since 1994.

2.1. PG morphophonology

PG is an agglutinative language; many different morphemes are expressed as prefixes or suffixes within the verbal complex (Ayala 1996; Hamidzadeh & Russell 2014; Zubizarreta 2022). The full verbal complex template is shown in (2):

(2) *Full verbal complex template in PG*

- | | |
|-------------------------------|---------------------------------------|
| a. optative prefix | t(V)- |
| b. negative prefix | n(V)- |
| c. agreement prefix | 1SG.A a-, 2SG.A re-, etc. |
| d. voice prefixes | AGD je-, RECIP jo-, CAUS mō- |
| e. (incorporated noun) | |
| f. epenthesis | j- |
| g. ROOT | |
| h. derivational suffixes | ATTEN -vĩ, TRANS -ka, etc. |
| i. control predicate suffixes | DESID -se, TOTAL -pa, etc. |
| j. magnitude suffixes | INTENS -ite, DIM -ʔi, etc. |
| k. mood suffixes | FUT -ta, FRUS -mōʔã, etc. |
| l. various other suffixes | ‘in vain’ -rei, ‘pretend’ -waʔu, etc. |
| m. negative enclitic | NEG =(r)i |
| n. various enclitics | Q =pa, COND =rãmō, etc. |

This order of elements in the verbal complex interacts with morphophonological processes in the language. The root and its prefixes form a prosodic word, which is the unit of nasalisation and stress assignment (Russell 2022b), while suffixes constitute separate prosodic words of their own (Dąbkowski 2022).

Table 1. PG pronoun paradigm.

	Full	A	B
ISG	ʃe	a-	ʃe=
IPL.INCL	ɲānde	ja-	ɲānde=
IPL.EXCL	ore	ro-	ore=
1>2SG	N/A	ro-	N/A
1>2PL	N/A	po-	N/A
2SG	nde	re-	nde=
2PL	pěě	pe-	pěende=
3	haʔe	o-	i=

2.1.1. Agreement. PG makes use of two sets of agreement morphology, which I term ‘Set A’ and ‘Set B’ here. In Table 1, I present the full independent form of each subject pronoun, followed by the corresponding Set A and B forms. Set A agreement has been previously analysed as true phi-agreement, as opposed to the Set B clitics (Woolford 2016; Zubizarreta & Pancheva 2017).

Set A agreement is used when agreement is controlled by the transitive subject (as in (3a)) or active intransitive subject (as in (3b)).

- (3) a. a-h-eʃa araʔi
 ISG.A-H-see cloud
 ‘I see a cloud’. 20210216_mcg
- b. a-jeroʔki
 ISG.A-dance
 ‘I dance’. 20201112_mcg

Set B forms are used when agreement is controlled by the transitive object (as in (4a)) or stative intransitive subject (as in (4b)). The system of argument cross-reference in PG is a case of hierarchical agreement: subject and object cross-referencing morphemes compete for a single slot, and the winner is the form higher on the person hierarchy (Velázquez-Castillo 1991; Nichols 1992; Siewierska 1996). Additionally, Set B forms are used in possessive contexts, as in (4c).

- (4) a. haʔe ʃe=ʊʷeʔru kaʔaʊʷi=pe
 3 ISG.B=bring forest=LOC
 ‘He brought me to the forest’. 20200929_mcg
- b. ʃe ʃe=r-oʔi
 ISG ISG.B=R-cold
 ‘I am cold’. 20200916_mcg
- c. ʃe=r-ouʃa
 ISG.B=R-house
 ‘my house’ 20201105_mcg

2.1.2. Stress. Primary stress is systematically assigned to the final syllable of the prosodic word. A root and its prefixes constitute a single prosodic word: a root-final syllable receives primary stress, as in (5), with a few lexical exceptions (Gregores & Suárez 1967).

- (5) a-ʊʷaʔta
 ISG.A-walk
 ‘I walk’. 20210401_ixo

Suffixes in PG fall into one of two domains, classified by whether the suffix can bear stress (Mistieri 2013). This is a property of the suffix itself: for instance, the totalitative suffix in (6a) is stress-bearing, while the future suffix in (6b) is not. As exemplified by the final syllable of the root ‘walk’ in (6a), the final syllable of an embedded prosodic word is realised with secondary stress, even when this results in a surface stress clash.

- (6) a. a-ʉ^wa,ta-‘pa
ISG.A-walk-TOTAL
‘I finished walking’. 20210503_mcg
- b. a-ʉ^wa’ta-ta
ISG.A-walk-FUT
‘I will walk’. 20210401_ixo

The ability to bear stress is not strictly conditioned by semantics: in fact, whether or not a suffix bears stress is actually attributable chiefly to historical origin. Stress-bearing suffixes can generally be reconstructed as having coda consonants in Proto-Tupí-Guaraní, while non-stress-bearing suffixes ended in open syllables (Jensen 1998; Russell 2023). All stress-bearing suffixes linearly precede all non-stress-bearing suffixes and enclitics, and primary stress is assigned to the final syllable within the domain of stress-bearing suffixes.

3. Nasality and nasalisation in Paraguayan Guaraní

3.1. Nasality

PG distinguishes six vowel qualities, as shown in Figure 1. Vowel length is not contrastive. All oral vowels have nasal counterparts.

Vowel nasality is contrastive only if the vowel is stressed (Barratt 1981; Walker 1999). Typically, the vowel which receives the primary stress within a lexical item is the root-final vowel (Adelaar 1994; Cabral & Rodrigues 2011; Mistieri 2013). Preceding a stressed root-final nasal vowel, all previous vowels within the word also predictably surface as nasal. I interpret this generalisation as evidence that only a vowel which receives stress may be specified phonologically as either oral or nasal: all other vowels within the phonological word are underlyingly unspecified for nasality, and predictably surface as either nasal or oral due to the influence of the stressed vowel. Therefore, the minimal pair [tu’pa] ‘bed’ vs. [tũ’pã] ‘God’ exists in the language, but there are no possible counterparts like *[tũ’pa] or *[tu’pã], in which nasality is contrastive on an unstressed, non-root-final vowel.

The inventory of consonants in PG, including those of ‘mixed’ articulation, is represented in Table 2. Notably, there are no voiced stops in the inventory. However, segments of ‘mixed’ articulation, in which a consonant begins as a nasal and ends as a voiced stop, are present in PG. The status of these ‘mixed’ segments has long been debated in the Tupí-Guaraní literature. Analyses vary between arguing that they are the surface result of the pre-nasalisation of oral stops (Gregores & Suárez 1967; Rose 2008; Daviet 2016) versus the post-oralisation of nasal consonants (Piggott 1992; Cardoso 2009; Lapierre & Michael 2018; Estigarribia 2021). I take the latter position, that these ‘mixed’ segments are post-oralised allophones of nasal consonants. The surface form of an underlying nasal consonant is determined by its

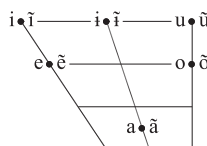


Figure 1. Paraguayan Guaraní phonemic vowel inventory.

Table 2. Paraguayan Guaraní consonant inventory, including phonemes and mixed-articulation allophones.

	Bilabial	Alveolar	Postalveolar	Palatal	Velar	Labial-velar	Glottal
Plosive	p	t			k	k ^w	ʔ
Nasal		m		ɲ	ŋ	ŋ ^w	
Mixed		mb			ŋg	ŋg ^w	
Flap							
Fricative		s	ʃ				h
Approximant	v			j	ɰ	ɰ ^w	
Lateral		l					

Table 3. Allophonic consonant alternations before oral vs. nasal vowels.

	Underlying	_V	_Ṽ
/N/	/m/	[mb]	[m]
	/n/	[nd]	[n]
	/ŋ/	[ŋg]	[ŋ]
	/ŋ ^w /	[ŋg ^w]	[ŋ ^w]
/J/	/ɾ/	[ɾ]	[ɾ̃]
	/v/	[v]	[ṽ]
	/j/	[j] ~ [j̃]	[j]
	/ɰ/	[ɰ]	[ɰ̃]
	/ɰ ^w /	[ɰ ^w]	[ɰ̃ ^w]

immediate phonological environment: before an oral vowel, a nasal stop surfaces as its partially oralised allophone, as the result of shielding (Stanton 2017). Before a nasal vowel, a nasal consonant surfaces faithfully as nasal, while an underlying oral approximant surfaces as nasalised. Consonants alternate as in Table 3.

3.2. Nasalisation

Both regressive (leftward) and progressive (rightward) spreads of nasalisation are attested in PG. I argue that these two types of nasalisation differ in important ways. In this study, I investigate what conditions variation in nasalisation, and how that variation is connected to the direction of nasalisation.

3.2.1. Regressive nasal harmony. Many languages of the Tupí–Guaraní family, including PG, exhibit long-distance regressive nasal harmony systems (Lapierre & Michael 2018; Miranda 2018; Baraúna 2020; Miranda & Picanço 2020). The harmony system in PG has been the subject of extensive description and analysis in the theoretical literature for decades (Gregores & Suárez 1967; Goldsmith 1976; Piggott 1992; Steriade 1993; Beckman 1998; Walker 1999, 2000; Kaiser 2008). Nasality spreads leftwards from a phonemic nasal vowel, and the domain of nasal harmony is the root and its prefixes (Lapierre & Michael 2018). The effects of nasal harmony are clear from the juxtaposition of (7a) and (7b). The two sentences differ in whether the root – ‘know’ in (7a) vs. ‘hug’ in (7b) – contains a phonemic nasal vowel, which in turn results in different surface forms for all the other morphemes within the word.

- (7) a. ndo-ro-jo-k^wa'a=j
 NEG-IPL.EXCL.A-RECIP-know=NEG
 'We don't know each other'. 20210316_mcg
- b. nō-řō-nō-ānū'ā=ř
 NEG-IPL.EXCL.A-RECIP-hug=NEG
 'We don't hug each other'. 20210315_ixo

The example in (7b) demonstrates that the presence of a phonemic nasal vowel results in nasalisation of segments to its left within the word. In PG, a phonemic nasal consonant additionally triggers regressive nasalisation, as in (8). I assume that the underlying form of the root 'listen' is /h-enu/, with a nasal consonant /n/ and oral vowel /u/: the underlying nasal consonant surfaces as post-oralised before the oral vowel, and spreads its nasality leftward. As a result, prefixes before a root containing a nasal consonant surface as nasalised, as in (8). All prefixes show predictable effects of regressive nasal harmony.

- (8) nō-řō-nō-h-ē'ndu=j
 NEG-IPL.EXCL.A-RECIP-H-listen=NEG
 'We don't listen to each other'. 20210224_ixo

The voiceless obstruents /p/, /t/, /k/, /h/ and /ʔ/ are transparent to regressive nasal harmony, as visible in the roots 'cut' (9a) and 'hit' (9b), while all other segments in the inventory show surface effects of nasalisation.² No consonants in the inventory block the spread of nasality.

- (9) a. āř-kĩ'tĩ i'vi'ra ki'se=pe
 ISG.A-cut wood knife=LOC
 'I cut the wood with a knife'. 20210426_mcg
- b. ō-nē-nū'pā
 3.A-AGD-hit
 'He was hit'. 20201112_mcg

Just as a nasal consonant in a root triggers nasalisation of segments to its left, so does a nasal consonant in a prefix, like the initial segment of the causative prefix in (10a). However, a nasal vowel or consonant in a suffix does not trigger the nasalisation of material earlier in the word. For instance, the frustrative suffix contains a final nasal vowel /ā/, which triggers nasalisation within the suffix itself but not beyond, as in (10b).³

- (10) a. haʔe ō-nē-mbo-vi,ʔa-k^wa'a i=fu'pe
 3 3.A-AGD-CAUS-happy-know 3.B=DOM
 'He knows how to make himself happy'. 20201020_ixo
- b. ʃe a-vi,ʔa-mō'ʔā
 ISG ISG.A-happy-FRUS
 'I was almost happy'. 20201008_mcg

The presence of a suffix with an initial alveolar flap and following nasal vowel exceptionally results in the spreading of nasality one segment to the left:⁴ in (11), for instance, the final vowel /o/ of the root

²There is one exception to this generalisation: the alveolar lateral, which is a voiced sonorant, does not show any effect of nasalisation and could therefore also be considered to be transparent to nasal harmony. However, this segment has a unique distribution in PG, as it is found only in ideophones and loanwords (Estigarribia 2020).

³I assume that the domain of nasal harmony in PG is the prosodic word, and that the frustrative suffix [-mōʔā] constitutes a prosodic word of its own. This analysis is essentially a reframing in prosodic terms of the proposal put forth by Beckman (1998) that stressed syllables, both primary and secondary, block nasal harmony.

⁴The surface nasalisation described in this paragraph applies in the case of the nominal future suffix /-rā/, shown in (11), and in the case of the conditional suffix /-ramō/, sometimes pronounced as [-řō].

‘wife’ surfaces as phonetically nasal, but does not trigger nasalisation of segments further to the left. I assume that this surface form is articulatorily, rather than phonologically, driven.

- (11) /je=remireko-rã/
 [jẽ=rẽmbirekõ-řã]
 ISG.B=wife-N.FUT
 ‘my fiancée’

In summary, nasality spreads from a phonemic nasal vowel or consonant within the phonological word in PG. This nasal harmony process applies from right to left, as it affects segments to the left of the trigger. Regressive nasal harmony targets all segments except voiceless obstruents, which are transparent to harmony.

3.2.2. *Progressive nasalisation.* Nasalisation in PG also has a limited rightward spread. Following a stressed nasal vowel, the locative enclitic /pe/ may undergo nasalisation to surface as [=mẽ], as illustrated in (12).

- (12) a. tu'pa=pe
 bed=LOC
 ‘to the bed’20210303_ixo
 b. tũ'pã=mẽ
 God=LOC
 ‘to God’

While regressive nasalisation spreads throughout roots and all prefixes, progressive nasalisation occurs only with a very select group of targets (Lunt 1973; Humbert & Piggott 1997). In his grammar of PG, Estigarribia (2020) lists 11 suffixes and enclitics affected by progressive nasalisation: each morpheme has one allomorph following an oral vowel, and a different (nasal) allomorph following an oral vowel; these are shown in Table 4. I find that only five of these 11 suffixes productively undergo nasalisation for the speakers with whom I worked, listed as ‘productive’ in Table 4.⁵ The productive morphemes include the totalitative (TOTAL), which indicates application and/or exhaustion of a given event to the whole object or subject (Estigarribia 2020); the locative (LOC); an enclitic indicating ‘until’; the incipient (INCIP), representing the immediate future; and the nominal past (N.PST), which restricts the temporal interpretation of a noun to the past, similar to English ‘former’ (Tonhauser 2007). The unproductive morphemes include the collective (COLL), indicating plurality of countable items; the collective plural (COLL.PL), indicating a place of abundance; various passive nominalisers; and the enclitic meaning ‘towards’.

While voiceless stops are transparent to regressive nasal harmony, progressive nasalisation exclusively targets morphemes with initial voiceless stops.⁶ Some suffixes and enclitics undergo total nasalisation – for example, LOC [pe] ~ [mẽ] – while the initial voiceless stop of others pre-nasalises – for example, TOTAL [pa] ~ [mba]. This distinction is attributable to historical origin (Russell 2023): suffixes which pre-nasalise have earlier origins as roots,⁷ and the pre-nasalisation of voiceless-stop-initial roots was a productive process in Proto-Tupí-Guaraní (Estigarribia 2021).

⁵It is unclear exactly what might account for the difference in the number of morphemes with nasal allomorphs that I present here vs. the number given in Estigarribia’s grammar (e.g., social factors like age or region). For the speakers with whom I have worked, the collective plural and passive nominaliser morphemes do not appear to be productive in any form, oral or nasal. The collective and ‘towards’ morphemes can productively attach, but always appear in their oral forms.

⁶The enclitic =yoto ‘towards’ begins with a velar approximant in PG; however, it is reconstructed in Proto-Tupí-Guaraní as *kotj with an initial voiceless stop (Jensen 1998: 514). As suggested by a reviewer, the fact that the initial consonant of this morpheme no longer forms a natural class with the others may provide part of the explanation for why it is synchronically a non-undergoer.

⁷The totalitative [pa] ~ [mba] is reconstructed as originating from the Proto-Tupí-Guaraní root *pab ‘to finish’ (PG pa; Jensen 1998: 537) and the incipient [pota] ~ [mbota] from the root *potár ‘to want’ (PG pota; Jensen 1998: 536).

Table 4. Allomorphs of target suffixes in Paraguayan Guaraní.

	V ₋	Ṽ ₋	Gloss
Productive	-pa	-mba	TOTAL
	=pe	=mẽ	LOC
	=peve	=mẽve	‘until’
	-pota	-mbota	INCIP
	-k ^w e	-ŋ ^w ẽ	N.PST
Unproductive	=k ^w era	=ŋg ^w era	COLL
	-ti	-ndi	COLL.PL
	-pi	-mbi	NMLZ.PASS
	-pire	-mbire	NMLZ.PASS.PST
	-pirã	-mbirã	NMLZ.PASS.FUT
	=uqoto	=ŋuqoto	‘towards’

While regressive nasalisation may be triggered by either a phonemic nasal vowel or a consonant, the same is not true for progressive nasalisation. If a nasal consonant is followed by an oral vowel, and therefore surfaces as post-oralised, as in the root /h-enu/ ‘listen’ in (13), nasality cannot spread rightwards, as shown in (13b). Therefore, only stressed nasal vowels should be able to trigger progressive nasalisation in PG.

- (13) a. ã-h-ẽ'ndú
 ISG.A-H-listen
 ‘I listened’.
- b. ã-h-ẽ.ndu-'pa *ãhẽndumba
 ISG.A-H-listen-TOTAL
 ‘I finished listening’.
- 20210223_ixo

Progressive nasalisation involves the rightward spread of nasality. The trigger is a phonemic nasal vowel, and the target is a suffix-initial syllable beginning with a voiceless stop belonging to one of the five productive suffixes listed in Table 4. The surface effect of progressive nasalisation differs based on the suffix: some suffixes show the effects of full nasalisation, while others pre-nasalise.

3.3. Interim summary

I have described two processes of nasalisation in PG. In Table 5, I compare several properties of these processes. Regressive harmony targets all preceding segments except for voiceless obstruents, which are transparent. Progressive nasalisation, on the other hand, exclusively targets syllables with initial voiceless stops. Additionally, regressive harmony appears to be fully productive, while progressive nasalisation is limited to a very small number of particular suffixes and enclitics, with morpheme-specific effects.

I now turn to the interactions of these two nasalisation processes with roots of Spanish etymological origin. Russell (2022a) finds that the presence of a nasal consonant in a Spanish-origin root can trigger the nasalisation of prefixes. The observation that regressive nasal harmony also applies to some extent to Spanish-origin roots provides support for the claim that it is in fact a productive process in the language. However, the interactions of progressive nasalisation and Spanish-origin roots have gone unreported in prior literature; I seek to fill that gap here. Gaining insight into these interactions is crucial to an understanding of the differences between regressive and progressive nasalisation in PG.

Table 5. Comparing directions of nasalisation in Paraguayan Guaraní.

	Regressive harmony	Progressive nasalisation
Direction	R-to-L	L-to-R
Trigger	Nasal V or C	Nasal V
Target	All segments except voiceless obstruents	Suffix-initial voiceless stop (within a handful of morphemes)
Morphological conditioning	None	Morpheme-specific (pre-nasalisation or full nasalisation)
Domain	Root + prefixes	Suffix-initial syllable

4. Spanish-origin items

Paraguay has been cited as a case of a country in which stable bilingualism is the norm (Romaine 1995; Trudgill 1995). According to the 2012 census of Paraguay, 77% of Paraguayans speak Guaraní (DGEEC 2012). Of that 77%, 8% do not speak Spanish. Additionally, 73% of Paraguayans speak Spanish, of which 5% do not speak Guaraní. Since the majority of the population is bilingual in PG and Spanish, there is no clear division between the two languages in everyday life. The term *Jopará*, which means ‘mixture’ in PG, refers to the colloquial variety which involves frequent language mixing between PG and Spanish (Lustig 2010; Estigarribia 2015). The intricacies of *Jopará* have been documented and described in many publications, including Morinigo (1959); Melià (1974); Boidin (2006); Bakker *et al.* (2008); Palacios Alcaine (2008); Dietrich (2010); Lustig (2010); Cardona (2008), among others. The use of morphemes from both Spanish and PG is extremely common at all levels of formality of language use in Paraguay, but the proportion contributed from each language varies considerably (Estigarribia 2015), and sociolinguistic work has focused on the comparative uses of PG and Spanish in different spheres (Gynan 1998, 2011; Choi 2003, 2004, 2005; Zajićová 2009). Studies of language attitudes in Paraguay have found that the population generally holds positive attitudes towards both languages, albeit for different reasons: people report a sense of pride and identity related to the use of PG, and connect the use of Spanish to economic value (Choi 2003; Gynan 2011). Language use in Paraguay is fundamentally multilingual, and speakers recruit their knowledge of both languages in everyday speech. In this study, I focus specifically on the use of lexical items of PG and Spanish origin in PG.

4.1. Phonotactic adaptation

Individual Spanish-origin lexical items display various repairs of violations of PG phonotactics. Pinta & Smith (2017) propose five lexical strata in PG, based on phonological repairs of loanwords from Spanish, as reproduced in Table 6. They identify four different properties: the presence of a nasal coda (N CODAS), any coda consonant (CODAS), non-final stress (NON-FINAL STRESS and complex onsets (#CC). The more properties are repaired, the more native-like a stratum is, and conversely, the more properties are tolerated, the more foreign-like a stratum is. Pinta and Smith do not consider behaviour in contexts of nasalisation in their analysis of lexical strata.

I provide an example of a lexical item from each stratum in Table 7; examples are taken from Pinta & Smith’s (2017) discussion of their proposed strata. In the nativ(ised) stratum, which includes native Guaraní items as well as fully nativised loans, all syllables are open, stress is final and no complex onsets occur, as visible from the adaptation of *culantro* ‘coriander’ to [kurã’tũ]. In the mostly nativised stratum, all syllables are open and stress is final, but onset consonant clusters are tolerated, as in the initial syllable of ‘Pluto’. In the partially nativised stratum, all syllables are open, but non-final stress is tolerated, and so are onset consonant clusters, as evidenced by the adaptation of ‘London’ with the onset cluster in the second syllable intact, but the final coda /s/ deleted. In the barely nativised stratum, non-nasal codas are tolerated, as are non-final stress patterns and onset consonant clusters; however,

Table 6. *Lexical strata in Paraguayan Guaraní; reproduced from Pinta & Smith (2017: 306).*

	N CODAS	CODAS	NON-FINAL STRESS	#CC
1. Nativ(ised)				
2. Mostly nativised	Repaired	Repaired	Repaired	<i>Tolerated</i>
3. Partially nativised	Repaired	Repaired	<i>Tolerated</i>	<i>Tolerated</i>
4. Barely nativised	Repaired	<i>Tolerated</i>	<i>Tolerated</i>	<i>Tolerated</i>
5. Unadapted	<i>Tolerated</i>	<i>Tolerated</i>	<i>Tolerated</i>	<i>Tolerated</i>

Table 7. *Examples of adaptations sorted by stratum in Paraguayan Guaraní.*

	Spanish		Guaraní	Gloss
1. Nativ(ised)	<i>culantro</i>	[ku. 'lan.tro]	[kurã 'tũ]	'coriander'
2. Mostly nativised	<i>Plutón</i>	[plu. 'ton]	[plu. 'tõ]	'Pluto'
3. Partially nativised	<i>Londres</i>	['lon.dres]	['lo.ndre]	'London'
4. Barely nativised	<i>almacén</i>	[al.ma. 'sen]	[al.ma. 'sẽ]	'department store'
5. Unadapted	<i>ensalada</i>	[en.sa. 'la.ða]	[en.sa. 'la.da]	'salad'

nasal codas are repaired, as visible from the adaptation of the nucleus and coda of 'department store' as a nasal vowel. Finally, in the unadapted stratum, non-final stress and onset consonant clusters are tolerated, and codas of all kinds are tolerated, including nasal ones, as in the first syllable of 'salad'.

The stratum into which a loan falls is partially a reflection of the time depth of the loan, à la Itô & Mester (1999), but is mainly dependent on the loan's phonotactic similarity to PG in its unadapted form. For instance, adaptation of the Spanish word *papá* 'father' into PG necessitates no overt repair of PG phonotactics, and would thereby be classified as Level 1 ('nativ(ised)'). I acknowledge the limits of this system of lexical strata, and use it here simply as a proxy measure for quantifying the well-formedness of a Spanish-origin lexical item vis-à-vis PG phonotactics.

4.2. Nasality

Paraguayan Spanish has three phonemic nasal consonants (bilabial /m/, alveolar /n/ and palatal /ɲ/), but no phonemic nasal vowels (Cassano 1971). It is not the case that vowels are never pronounced with phonetic nasalisation in Spanish; however, nasalisation is not a target of speech production in Spanish vowels, and nasalisation is not a phonologically active feature of vowels in the language (Solé 1992). Spanish borrowings contain surface phonotactic environments that never occur natively in PG, since phonemic nasal consonants in PG always surface as partially oralised allophones before oral vowels. An underlying sequence /ma/ must be realised either as [mba] or as [mã] on the surface in PG due to phonotactic constraints in the language; however, lexical items of Spanish origin are not subject to the same phonotactic constraints. In some older borrowings, this mismatch in phonotactics is in fact repaired, as shown in Table 8. Some sequences of a vowel and nasal coda consonant in Spanish borrowings are reinterpreted in PG as phonemic nasal vowels. In words like 'heart', 'pants' and 'soap', a word-final VN sequence in Spanish is reinterpreted as a stressed nasal vowel in PG, and these words obey nasal harmony within the root. The word for 'pillow' demonstrates the outcome when a nasal consonant in a Spanish-origin root does not appear within the same syllable as the stressed vowel: this /m/ spreads its nasality leftwards, and surfaces as post-oralised [mb], exactly as expected given the phonotactic constraints of PG.

However, most Spanish-origin items are pronounced roughly as they are in Spanish: nasal–oral sequences are not repaired, and these items do not exhibit root-internal nasal harmony. The difference

Table 8. Adaptations of selected Spanish-origin items with nasal consonants into Paraguayan Guaraní.

Spanish		PG	Gloss
<i>corazón</i>	kora'son	kõrã'sõ	'heart'
<i>calzón</i>	kal'son	kã'sõ	'pants'
<i>jabón</i>	xa'βon	hã'ũõ	'soap'
<i>almohada</i>	almo'aða	ãrãmbo'ha	'pillow'

in strategy between adaptations to morpheme-internal nasal harmony and apparent non-adaptations is likely related to the diachronic expansion of PG–Spanish bilingualism in Paraguay. Spanish influence on PG took place slowly and gradually over a long period of time until quite recently, when urbanisation has accelerated the rate of Guaraní speakers acquiring and using Spanish (Zajićová 2009; Fernández Barrera 2015). Since the majority of users of PG today are bilingual with Spanish, they seem to be more tolerant of violations of PG phonotactics by Spanish-origin items (Pinta 2013; Pinta & Smith 2017). Earlier borrowings, therefore, underwent nativisation more than recent borrowings, which maintain the phonology of Spanish more faithfully.

4.3. Nasalisation

The presence of a nasal consonant in a Spanish-origin root triggers nasalisation of prefixes (Thun 2005; Russell 2022a), even when the vowels and consonants in between are oral, as illustrated in (14). Examples below are presented using a four-line method, in which the top line is the assumed underlying representation, with Spanish-origin roots (in Spanish orthography) italicised, and underlying nasal consonants bolded. In the second line – the surface IPA representation – the targets of nasalisation are underlined.

- (14) a. /ja-jo-traiciona-ta/
 [nã-jõ-traisjo 'na-ta]
 IPL.INCL.A-RECIP-betray-FUT
 'We are going to betray each other'. 20210329_mcg
- b. /nV-ja-je-reconoce=i/
 [nã-jã-jẽ-rekono 'se=j]
 NEG-IPL.INCL.A-AGD-recognise=NEG
 'We don't recognise ourselves'.
- c. /n-o-je-solucion=i=va/
 [n-õ-jẽ-solusjo 'na=j=va]
 NEG-3.A-AGD-solve=NEG=REL
 'that it would not be solved'

The presence of a nasal consonant may also trigger nasalisation in suffixes, even when the vowels and consonants in between are oral, as in (15).

- (15) a. /ʃe a-ĩ cocina=pe/
 [ʃe ã'ĩ kosi 'na=mẽ]
 ISG ISG.A-be kitchen=LOC
 'I am in the kitchen'. 20210217_ixo

- b. /o-je-*maquilla*=peve ʃe n-a-*reconoce*=i ʃupe/
 [õ-jẽ-*maki* 'la=*mëve* ʃe n-ã-*rekono* 'se=j ʃu'pe]
 3.A-AGD-makeup=until ISG NEG-ISG.A-recognise=NEG DOM
 'Until she put on makeup, I didn't recognise her'. 20210303_ixo
- c. /a-*nada*-pa=mã/
 [ã-na, ða-'*mba*=mã]
 ISG.A-SWIM-TOTAL=already
 'She finished swimming'. 20210426_mcg

However, importantly, both consultants in an elicitation setting often expressed uncertainty regarding the acceptability of the combination of a Spanish-origin root and suffix nasalisation, as illustrated in (16).

- (16) a. /ʃe a-*soña*-pa-ite=mã/
 [ʃe a-so ʃa-, pa-j' te=mã]
 ISG ISG.A-dream-TOTAL-INTENS=already
 'I finished dreaming'. 20210421_ixo
- b. /ʃe a-*soña*-pa-ite=mã/
 [ʃe a-so ʃa-, *mba*-j' te=mã]
 ISG ISG.A-dream-TOTAL-INTENS=already
Elicitor: Can I also say this to mean 'I finished dreaming'?
Consultant: Yes, some people say it, but I'm not sure which is correct (*pa* or *mba*).

Spanish-origin roots may participate in nasalisation of prefixes and suffixes, but not in root-internal harmony, with the exception of the few nativised borrowings. The interactions of Spanish-origin items with regressive nasal harmony have previously been analysed as the innovation of a novel system of consonant harmony, in which a nasal consonant in a Spanish-origin root is in correspondence with consonants in prefixes, resulting in nasalisation of those prefixal consonants (Russell 2022a). Interactions of Spanish-origin roots and progressive nasalisation are still unclear at this point. Although both consultants produce and accept some forms in which a nasal consonant in a Spanish-origin root triggers progressive nasalisation in an elicitation setting, they also verbally express uncertainty about the acceptability of similar forms.

5. Quantifying variation

In the elicitation context in which data were collected, variation in actual application of nasalisation abounded. In separate elicitation sessions, a consultant provided two different PG forms for the same English sentence: one in which no prefixes nasalised, in (17a), and one in which only the prefix closest to the root nasalised, in (17b). The same consultant also accepted a third form, shown in (17c), in which all prefixes were pronounced as nasalised, in the second session.

- (17) /no-ro-jo-*reconoce*=i/
 a. [ndo-ro-jo-*rekono* 'se=j]
 NEG-IPL.EXCL.A-RECIP-recognise=NEG
 'We don't recognise each other'. 20210316_mcg
- b. [ndo-ro-*nõ*-*rekono* 'se=j]
 NEG-IPL.EXCL.A-RECIP-recognise=NEG
 'We don't recognise each other'. 20210329_mcg

c. [nõ-rõ-nõ-rekono 'se=j]

NEG-IPL.EXCL.A-RECIP-recognise=NEG

Elicitor: Can I also say this to mean 'we don't recognise each other'?*Consultant:* Yes, you can say that too.

The quantitative analysis of variation is not informative in an elicitation context, because the frequency of forms is determined by the question-and-answer format of the elicitation interaction and may not reflect the distribution of forms in spontaneous language use. I, therefore, now turn to corpus data from sociolinguistic interviews to examine variation in nasalisation in PG and its interactions with direction of nasalisation and root etymological origin.

5.1. Methods

The data presented here come from a corpus of sociolinguistic interviews with native PG speakers in Paraguay (Bittar Prieto 2021). Twenty-six interviews were conducted; each interview lasted roughly an hour. Half of the interview data was collected in the urban Bañado Sur neighbourhood of Asunción in 2015. These conversational interviews were conducted by Israel Pedrozo, a resident of Bañado Sur and native speaker of PG and Paraguayan Spanish. The age of the participants in the urban community ranged from 18 to 68 (mean = 43.9, *SD* = 18.7). The other half of the data was collected in the rural community of San Juan Nepomuceno, roughly 200 kilometres from Asunción. The interviews in this community were conducted between October 2019 and January 2020 by Antonio Zena, a native of San Juan Nepomuceno. The ages of the participants in the rural area ranged from 18 to 73 (mean = 45.6, *SD* = 15.7). For each interview, all utterances were transcribed in PG and translated to Spanish by the interviewers themselves as well as by Josefina Bittar, a linguist and heritage speaker of PG.

All potential sites of nasalisation – both instances in which nasalisation of an affix actually occurred and those in which it could have but did not – were added to a data frame. The relevant variable context is the combination of an affix affected by nasalisation and a nasal root: either a PG-origin root containing a phonemic nasal segment, or a Spanish-origin root containing a nasal consonant. Data collection was limited to cases of certain affixes, specifically those affixes in which consonants clearly alternate between oral and nasal, listed in Table 9. Nasalisation of affixes that comprise only a single vowel, or a single vowel and an obstruent, was not included in the data set, due to the limitations of measuring nasality in these contexts without data about nasal airflow. Additionally, tokens were excluded when it was not possible to ascertain the etymological origin of the root (e.g., *mamá* 'mother').

Though the nominal past tense suffix *k^we* had been found to productively nasalise in the elicitation context, there were no tokens of its nasal form in the corpus, so it was excluded from the data set. The corpus was also checked for the six target suffixes which had not been found to productively nasalise in the elicitation context (cf. Table 4): the collective *k^wéra*; the collective plural *tĩ*; the passive nominalisers *pĩ*, *pire* and *pirã*; and the enclitic 'towards' *uoto*. The corpus included two tokens of the nasal allomorph of the collective *k^wéra*, both following the lexical item *mĩtã* 'child'. These tokens were not included in the final data set. The corpus included no tokens of the collective plural or any of the passive nominalisers, and no tokens of the nasal allomorph of 'towards' *uoto*.

The total number of tokens included in the data set was 3,641. Each token, defined as an affix that is a potential undergoer of nasalisation, was coded for the dependent variable – viz., whether nasalisation of an affix occurred – as well as a number of linguistic and social factors. Social factors included gender (self-identified as male or female), age and community affiliation (rural or urban). Linguistic factors included direction of nasalisation (regressive or progressive), morpheme identity of the target affix, etymological origin of the root (PG or Spanish), morphological category (adjective, noun or verb) and log-transformed root token frequency within the corpus.⁸ Gender, age and community affiliation (rural

⁸Log-transformed frequency was used in order to dampen potential effects of a small number of very high-frequency roots. An alternative formulation of frequency was also coded. In this formulation, the number of times the root triggering nasalisation

Table 9. *Targets of nasalisation included in the data set.*

	Gloss	Oral	Nasal
Before root	IPL.INCL.A	ja	ɲã
	IPL.INCL.B	ɲãnde	ɲãnẽ
	2PL.B	pẽnde	pẽnẽ
	2SG.B	nde	nẽ
	AGD	je	ɲẽ
	CAUS	mbo	mõ
	J	j	ɲ
	NEG	nd(V)	n(ĩ)
	RECIP	jo	ɲõ
After root	INCIP	pota	mbota
	LOC	pe	mẽ
	TOTAL	pa	mba
	‘until’	peve	mẽve

or urban) were included in order to investigate whether social factors have any significant impact on application of nasalisation. Older people and rural communities are often associated with a form of PG which is considered more ‘pure’, while young people and urban communities may be associated with bilingualism and language mixing (Rubin 1968; Gómez-Rendón 2007). The direction of nasalisation was identified as either regressive (leftward) or progressive (rightward). I have argued that the two types of nasalisation differ in important ways, and therefore hypothesise that they will show significantly different patterns. Several characteristics of the root were identified and included in the analysis: etymological origin (PG or Spanish), morphological category and log-transformed token frequency.

Tokens were also coded for three additional factors which are relevant only for Spanish-origin items: phonotactic well-formedness of the root with respect to PG phonotactics (measured from 1 to 5 as per Pinta & Smith’s classification), distance (in segments) between the trigger and target, and whether or not the nasal consonant trigger appears in the stressed syllable. Phonotactic well-formedness (lexical stratum) is relevant only for Spanish-origin items, as all PG-origin items in the corpus belong to the native stratum. If the phonological similarity of a Spanish-origin root to native PG items is a factor in predicting its rate of nasalising affixes, lexical stratum is expected to have a significant effect. The distance between the trigger and target was measured in terms of number of intervening segments, in order to investigate if distance has a significant impact on application of nasalisation. Since oral consonants and vowels may intervene between a trigger in a Spanish-origin root and a target, this distance ranges in the corpus from zero to ten segments. In every case of a PG-origin item triggering nasalisation, though, the distance is zero. Finally, stress was included as a factor in order to assess if the relationship between stress and nasality in PG holds for Spanish-origin items as well: namely, that nasality is contrastive only on a stressed vowel, as described in §3.1. Stress was coded as binary: Yes, if the nasal consonant trigger in a root appears within the stressed syllable, and No if the nasal consonant trigger is in an unstressed syllable. Because of PG-internal phonotactics, regressive nasal harmony can be triggered by either a nasal consonant or a stressed nasal vowel. If a nasal consonant in a Spanish-origin root behaves like a phonemic nasal consonant in a PG-origin root, it is predicted to be able to trigger regressive nasal harmony. However, progressive nasalisation may only be triggered by a stressed nasal vowel, never by a phonemic nasal consonant (cf. (13)). Therefore, every trigger of suffix nasalisation is predicted to be a nasal vowel in a stressed syllable. This factor was included in

occurs with a given affix was divided by total root token frequency within the corpus, in order to reflect the frequency of the affix+root collocation. Due to the potential overlap with root token frequency, the two factors were not included in the same model. Collocation frequency was not found to be an improvement over root token frequency in any model.

Table 10. *Distribution of tokens by direction and etymological origin.*

Direction	With PG-origin root	% Nasal	With SP-origin root	% Nasal	Total % nasal
Regressive	1,575	90.7%	1,172	77.6%	84.1%
Progressive	389	90.2%	505	5.7%	47.9%

order to evaluate whether a root in which a nasal consonant trigger appears within a stressed syllable nasalises suffixes at a significantly higher rate than those in which the trigger of nasalisation appears in an unstressed syllable.

5.2. Findings

As indicated in Table 10, the corpus includes more tokens of regressive nasal harmony than progressive nasalisation. Tokens of PG-origin triggers of nasal harmony slightly outnumber Spanish-origin triggers. However, Spanish-origin roots appear in contexts of progressive nasalisation more frequently than PG-origin roots. This is likely attributable to the high frequency of suffixes attaching to proper nouns like place names, which are often of Spanish origin. The rate of nasalisation triggered by PG-origin roots is over 90% for both directions. With Spanish-origin roots, however, we see a drastic difference in nasalisation rate between regressive harmony, where nasalisation occurs 77.6% of the time, and progressive nasalisation, where it occurs with a mere 5.7% of the tokens. It is clear that a more in-depth investigation of the difference in nasalisation rates along the axis of direction, as well as by etymological origin of the root, is necessary.

The nasalisation rates for specific affixes, shown in Table 11, prove to be particularly enlightening.⁹ A PG-origin root triggers nasalisation of most prefixes over 90% of the time, with three notable exceptions, namely the three Set B morphemes (i.e., the forms used when the transitive object or stative intransitive subject controls agreement). The lower nasalisation rate of these morphemes may reflect their distinct morphological status, as they have been analysed as proclitics (Woolford 2016; Zubizarreta & Pancheva 2017). The rate of prefix nasalisation triggered by Spanish-origin roots is generally slightly lower, but consistently above 85%. With the three Set B morphemes, however, nasalisation rates are again much lower. When it comes to suffixes and enclitics, we see differences between individual morphemes. The large gap between the relatively high rates of nasalisation following a PG-origin root and the quite low rates of nasalisation following a Spanish-origin root is apparent.

5.2.1. Regressive nasal harmony. The total number of tokens included in the data set of regressive harmony was 2,747 (1,575 tokens preceding roots of PG origin, plus 1,172 tokens preceding roots of Spanish origin). The variation in regressive nasal harmony was statistically modelled using mixed-effects logistic regression.¹⁰ The model included seven fixed effects – gender, age, community affiliation, target affix identity, root etymological origin, morphological category and frequency – and a by-speaker random intercept. The three factors specified in §5.1 that are relevant only for Spanish-origin items (lexical stratum of the root, distance between trigger and target and whether or not the nasal consonant trigger appears in the stressed syllable) were not included in this model due to the asymmetry in the data. The model also included an interaction term between the target of harmony and the etymological origin of the root. None of the social factors significantly improved model fit at the threshold of $p < 0.05$. Additionally, morphological category had no significant impact on model fit. Though root origin was not significant as a main effect, it is significant in its interactions with target identity. The non-significant predictors were removed, with the exception of those involved

⁹The high number of tokens of AGD before Spanish-origin roots is likely due to calquing from Spanish middle constructions with *se* (Bittar Prieto 2021).

¹⁰Logistic regressions were run using the lme4 package in R (Bates *et al.* 2015). Values for r^2 were calculated using the MuMIn package in R (Bartón 2022).

Table 11. *Distribution of tokens by affix and root etymological origin.*

	Affix	With PG- origin root	% Nasal	With SP origin root	% Nasal	Total % nasal
Before root	NEG	525	94.5	207	85.0	91.7
	IPL.INCL.A	148	97.3	54	88.9	95.0
	IPL.INCL.B	56	33.9	67	5.97	18.7
	2PL.B	23	30.4	17	17.6	25.0
	2SG.B	333	85.6	182	43.4	70.5
	AGD	290	97.9	434	91.0	93.7
	CAUS	108	93.5	29	100	94.9
	RECIP	36	100	37	91.9	95.9
	J	56	100	145	97.9	98.5
After root	INCIP	5	60	2	0	42.8
	TOTAL	180	99.4	31	45.2	90.4
	LOC	188	86.2	457	2.6	28.5
	‘until’	15	40	9	11.1	29.2

in significant interactions, resulting in the final model in Table 12. The marginal r^2 of this model is 0.348, representing the variance explained solely by the fixed effects, and the conditional r^2 is 0.363, representing the variance explained by the entire model. In the model summary, a positive estimate indicates that the given factor level favours the nasal variant compared to the baseline, while a negative estimate indicates that the level favours the oral variant. Plots are provided in Figures B1 and B2 in Appendix B.

Set B morphemes nasalise at a significantly lower rate than other morphemes. Root token frequency is significant in the model of regressive nasal harmony: the more frequent a root within the corpus, the higher the predicted rate of nasalisation. Etymological origin of the root is significant only in its interaction with two of the three Set B morphemes.

Variation in regressive nasal harmony triggered by only PG-origin items was modelled using mixed-effects logistic regression to isolate the factors relevant for those items. The total number of tokens included in this data set was 1,575. This model included six fixed effects – gender, age, community affiliation, target identity, morphological category and root token frequency – as well as a by-speaker random intercept. Again, none of the social factors significantly improved model fit at the threshold of $p < 0.05$. Morphological category and root frequency also had no significant impact on model fit. The non-significant predictors were removed, resulting in the final model in Table 13. The marginal r^2 of this model is 0.233, and the conditional r^2 is 0.240. The only significant predictor was the target identity of the morpheme: specifically, Set B morphemes nasalise at a significantly lower rate than other morphemes.

Variation in regressive nasal harmony triggered by only Spanish-origin items was modelled using mixed-effects logistic regression to isolate the factors relevant for those items. The total number of tokens included in this data set was 1,172. This model included nine fixed effects – gender, age, community affiliation, target identity, morphological category, frequency and two additional factors: root lexical stratum and distance – as well as a by-speaker random intercept, and an interaction between the target of harmony and the lexical stratum of the root. Again, none of the social factors significantly improved model fit at the threshold of $p < 0.05$. Morphological category also had no significant impact on model fit. The lexical stratum of the root was not significant, either as a main effect or in an interaction with the target morpheme identity. The non-significant predictors were removed, resulting in the final model in Table 14. The marginal r^2 of this model is 0.398, and the conditional r^2 is 0.422. Plots are provided in Figure B3 in Appendix B.

Table 12. Summary of model of regressive harmony.

	Estimate	Std. error	z-value	p-value	Sig.
(Intercept)	9.190×10^{-1}	3.174×10^{-2}	28.956	<0.001	***
MorphemeIPL.INCL.B (vs. MorphemeIPL.INCL.A)	-6.410×10^{-1}	4.492×10^{-2}	-14.268	<0.001	***
Morpheme2PL.B	-6.758×10^{-1}	6.455×10^{-2}	-10.470	<0.001	***
Morpheme2SG.B	-1.158×10^{-1}	2.918×10^{-2}	-3.967	<0.001	***
MorphemeAGD	9.297×10^{-3}	2.898×10^{-2}	0.321	0.748	
MorphemeCAUS	-3.572×10^{-2}	3.624×10^{-2}	-0.986	0.324	
MorphemeJ	2.368×10^{-2}	4.508×10^{-2}	0.525	0.599	
MorphemeNEG	-3.849×10^{-2}	2.678×10^{-2}	-1.437	0.151	
MorphemeRECIP	3.976×10^{-2}	5.334×10^{-2}	0.745	0.456	
RootOriginSpanish (vs. RootOriginGuarani)	-6.196×10^{-2}	4.600×10^{-2}	-1.347	0.178	
RootLogFrequency	1.081×10^{-2}	3.657×10^{-3}	2.957	0.003	**
MorphemeIPL.INCL.B:RootOriginSpanish	-1.797×10^{-1}	6.903×10^{-2}	-2.603	0.009	**
Morpheme2PL.B:RootOriginSpanish	-5.860×10^{-2}	1.018×10^{-1}	-0.576	0.565	
Morpheme2SG.B:RootOriginSpanish	-3.386×10^{-1}	5.264×10^{-2}	-6.433	<0.001	***
MorphemeAGD:RootOriginSpanish	1.398×10^{-2}	5.035×10^{-2}	0.278	0.781	
MorphemeCAUS:RootOriginSpanish	1.451×10^{-1}	7.509×10^{-2}	1.932	0.053	.
MorphemeJ:RootOriginSpanish	6.908×10^{-2}	6.405×10^{-2}	1.078	0.281	
MorphemeNEG:RootOriginSpanish	-5.729×10^{-3}	5.126×10^{-2}	-0.112	0.911	
MorphemeRECIP:RootOriginSpanish	-8.920×10^{-3}	8.105×10^{-2}	-0.110	0.912	

Table 13. Summary of model of regressive harmony triggered by PG-origin items.

	Estimate	Std. error	z-value	p-value	Sig.
(Intercept)	9.752×10^{-1}	2.160×10^{-2}	45.139	<0.001	***
MorphemeIPL.INCL.B (vs. MorphemeIPL.INCL.A)	-6.388×10^{-1}	4.011×10^{-2}	-15.926	<0.001	***
Morpheme2PL.B	-6.777×10^{-1}	4.011×10^{-2}	-11.731	<0.001	***
Morpheme2SG.B	-1.175×10^{-1}	2.598×10^{-2}	-4.522	<0.001	***
MorphemeAGD	4.084×10^{-3}	2.586×10^{-2}	0.158	0.875	
MorphemeCAUS	-3.974×10^{-2}	3.228×10^{-2}	-1.231	0.219	
MorphemeJ	2.270×10^{-2}	4.026×10^{-2}	0.564	0.573	
MorphemeNEG	-3.189×10^{-2}	2.390×10^{-2}	-1.334	0.182	
MorphemeRECIP	2.124×10^{-2}	4.748×10^{-2}	0.447	0.655	

In summary, Set B agreement morphemes nasalise at a significantly lower rate than all other morphemes. The token frequency of the trigger of nasal harmony significantly positively correlates with nasalisation rate. Spanish-origin roots do *not* trigger nasalisation at a significantly lower rate than PG-origin roots; however, root etymological origin is significant in its interaction with certain Set B morphemes. In such cases, Spanish-origin roots trigger nasalisation of Set B morphemes at a lower rate. When looking specifically at Spanish-origin roots, the lexical stratum of the root does not significantly

Table 14. Summary of model of regressive harmony triggered by Spanish-origin items.

	Estimate	Std. error	z-value	p-value	Sig.
(Intercept)	9.167×10^{-1}	5.216×10^{-2}	17.576	<0.001	***
MorphemeIPL.INCL.B (vs. MorphemeIPL.INCL.A)	-7.999×10^{-1}	5.845×10^{-2}	-13.684	<0.001	***
Morpheme2PL.B	-7.180×10^{-1}	8.965×10^{-2}	-8.009	<0.001	***
Morpheme2SG.B	-4.545×10^{-1}	5.054×10^{-2}	-8.991	<0.001	***
MorphemeAGD	2.534×10^{-2}	4.626×10^{-2}	0.548	0.584	
MorphemeCAUS	1.118×10^{-1}	7.345×10^{-2}	1.522	0.128	
MorphemeJ	8.135×10^{-2}	5.121×10^{-2}	1.589	0.112	
MorphemeNEG	-4.065×10^{-2}	4.887×10^{-2}	-0.832	0.406	
MorphemeRECIP	-3.075×10^{-2}	6.889×10^{-2}	-0.446	0.655	
RootLogFrequency	1.274×10^{-2}	6.103×10^{-3}	2.087	0.037	*
Distance	-2.518×10^{-2}	5.306×10^{-3}	-4.746	<0.001	***

Table 15. Summary of model of progressive nasalisation.

	Estimate	Std. error	t-value	p-value	Sig.
(Intercept)	0.86487	0.01985	43.561	<0.001	***
MorphemeTOTAL (vs. MorphemeLOC)	0.13096	0.02408	5.440	<0.001	***
MorphemeUNTIL	-0.47385	0.06122	-7.740	<0.001	***
RootOriginSpanish (vs. RootOriginGuaraní)	-0.83633	0.01977	-42.300	<0.001	***
MorphemeTOTAL:RootOriginSpanish	0.29073	0.04873	5.966	<0.001	***
MorphemeUNTIL:RootOriginSpanish	0.54703	0.09779	5.594	<0.001	***

affect the rate of nasalisation. An increase in distance between trigger and target significantly negatively correlates with nasalisation rate.

5.2.2. Progressive nasalisation. In statistical analysis of the progressive nasalisation data, the INCIP suffix was excluded from the model due to the low number of available tokens ($n = 7$). The total number of tokens included in the data set of progressive nasalisation was 880. The variation in progressive nasalisation was statistically modelled using mixed-effects logistic regression, including seven fixed effects – gender, age, community affiliation, target affix identity, root etymological origin, morphological category and frequency – and a by-speaker random intercept. The model also included an interaction term between the target morpheme identity and the etymological origin of the root. Again, none of the social factors significantly improved model fit. Additionally, neither the morphological category of the root nor root token frequency had a significant impact on model fit. The non-significant predictors were removed, resulting in the final model in Table 15. The marginal r^2 of this model is 0.779 and the conditional r^2 is 0.793. A plot is provided in Figure B4 in Appendix B.

Table 16. Summary of model of progressive nasalisation with PG-origin roots.

	Estimate	Std. error	<i>t</i> -value	<i>p</i> -value	Sig.
(Intercept)	0.882009	0.055569	15.872	<0.001	***
MorphemeTOTAL (vs. MorphemeLOC)	0.113746	0.080087	1.420	0.15637	
MorphemeUNTIL	0.732358	0.265141	2.762	0.00603	**
Frequency	−0.004157	0.010667	−0.390	0.69700	
MorphemeTOTAL:Frequency	0.003841	0.015463	0.248	0.80399	
MorphemeUNTIL:Frequency	−0.209630	0.044454	−4.716	<0.001	***

All three morphemes included in this model display distinctly different patterns of nasalisation. Spanish-origin roots are predicted to trigger nasalisation at a lower rate than PG-origin roots across the board; however, the size of this effect is specific to each morpheme.

Variation in progressive nasal harmony triggered by only PG-origin items was modelled using mixed-effects logistic regression to isolate the factors relevant for those items. The total number of tokens included in this data set was 389. This model included six fixed effects – gender, age, community affiliation, target identity, morphological category and root token frequency – as well as a by-speaker random intercept and an interaction term between target identity and root frequency. Again, none of the social factors significantly improved model fit at the threshold of $p < 0.05$. Morphological category also had no significant impact on model fit. The non-significant predictors were removed, resulting in the final model in Table 16. The marginal r^2 of this model is 0.228, and the conditional r^2 is 0.300.

When including only tokens of progressive nasalisation following a Spanish-origin root, the total number of tokens in this data set was 497. This model included nine fixed effects – gender, age, community affiliation, target affix identity, morphological category, frequency, root lexical stratum, stress and distance – and a by-speaker random intercept, as well as an interaction term between lexical stratum and affix morpheme identity. Again, none of the social factors significantly improved model fit, and neither the morphological category of the root nor frequency had a significant impact on model fit. The non-significant predictors were removed, resulting in the final model in Table 17. (Plots are provided in Figures B5 and B6 in Appendix B.) The marginal r^2 of this model is 0.376, and the conditional r^2 is 0.437. The probability of the nasal variant is predicted to be above zero only for those lexical items which most closely resemble PG-origin items in terms of lexical stratum, distance between target and trigger of nasalisation, and stress. Specifically, the Spanish-origin items which most closely resemble PG-origin items are those which include few, if any, violations of PG phonotactics, bear stress on a syllable containing a nasal consonant, and have little to no intervening material between that syllable and a prefix.

Looking at all contexts of progressive nasalisation, all morphemes nasalise at significantly different rates from one another, and Spanish-origin roots undergo nasalisation at a significantly lower rate than PG-origin roots. The interaction between root origin and morpheme identity is significant for all morphemes. When we consider only nasalisation triggered by Spanish-origin items, lexical stratum has a significant impact: the more phonotactically well-formed a root is in PG, the higher the rate of nasalisation. In fact, only those Spanish-origin roots which share their phonological properties with PG-origin roots are predicted to trigger progressive nasalisation at all. Roots in which the trigger of nasalisation appears within a stressed syllable nasalise at a significantly higher rate than those in which the trigger is in an unstressed syllable, signifying that the language-specific connection between stress and nasality in PG is maintained for Spanish-origin roots as well. An increase in distance between trigger and target significantly negatively correlates with nasalisation rate.

Table 17. Summary of model of progressive nasalisation triggered by Spanish-origin items.

	Estimate	Std. error	<i>t</i> -value	<i>p</i> -value	Sig.
(Intercept)	0.309904	0.039670	7.812	<0.001	***
Morpheme _{TOTAL} (vs. Morpheme _{LOC})	0.631322	0.081444	7.752	<0.001	***
Morpheme _{UNTIL}	0.252467	0.224856	1.123	0.262	
Stratum	−0.084342	0.010163	−8.299	<0.001	***
Stress _Y (vs. Stress _N)	0.112010	0.018720	5.983	<0.001	***
Distance	−0.030308	0.005867	−5.166	<0.001	***
Morpheme _{TOTAL} :Stratum	−0.211826	0.041799	−5.068	<0.001	***
Morpheme _{UNTIL} :Stratum	−0.064950	0.072044	−0.902	0.368	

6. Discussion

All nasalisation effects cannot be attributed to a single (morpho-)phonological process in PG. Findings from both data collected through elicitation and statistical modelling of a corpus of sociolinguistic interviews provide strong support for an analysis in which regressive nasal harmony and progressive nasalisation are in fact distinct processes, rather than reflections of bidirectional harmony. Regressive and progressive nasalisation differ in their effects and domains. While regressive nasalisation affects all segments except voiceless obstruents, progressive nasalisation targets only suffix-initial syllables beginning in voiceless stops, and the actual realisation of nasalisation depends on the individual morpheme. Regressive nasalisation is productive, to the point that even a nasal consonant within a Spanish-origin root may trigger nasalisation of prefixes. Nevertheless, this nasalisation pattern is distinct from that triggered by PG-origin roots, because it can operate at a distance. Progressive nasalisation, on the other hand, productively applies only for a very small number of suffixes and enclitics. Even among that limited set, actual rates of nasalisation are specific to each morpheme. Variation is additionally conditioned by different factors for each type of nasalisation. The rate at which regressive nasal harmony applies is predicted to differ significantly between different classes in the morphology, as proclitics and prefixes pattern distinctly from each other. This relationship between harmony and the morphosyntax points to a synchronically active process in which phonological constraints interact differently with proclitics and prefixes. In progressive nasalisation, however, phonologically or morphologically similar morphemes do not pattern together; rather, each morpheme behaves distinctly.

Additionally, PG regressive nasal harmony has been extended to Spanish-origin items: neither etymological origin nor lexical stratum is significant as a predictor of the application rate of nasal harmony. Even though Spanish-origin roots do not participate in root-internal harmony, a nasal consonant within such a root can trigger the nasalisation of prefixes and proclitics, constituting a case of innovated long-distance consonant harmony (Russell 2022a). Progressive nasalisation, though, has not been extended to Spanish-origin roots. Root etymological origin – PG vs. Spanish – accounts for the majority of the variation in progressive nasalisation. The rate of progressive nasalisation triggered by a Spanish-origin root is predicted to be above zero only when that Spanish-origin root closely resembles a PG-origin root in terms of phonotactics: lexical stratum, co-occurrence of stress and nasality on the same syllable, and distance between trigger and target of nasalisation are all significant. Interactions with root etymological origin point to the synchronic productivity of regressive nasal harmony, which is productive with all Spanish-origin items, as opposed to progressive nasalisation, which is limited only

to the most nativised of Spanish-origin lexical items. The findings contribute to the larger literature concerning the adaptation of loaned material to harmony systems. Cross-linguistically, it has been claimed that harmony applies at a lower rate, if at all, to loanwords (Clements & Sezer 1982; Ringen & Heinämäki 1999; Kertész 2003; Puthaval 2013). In this study, I have shown that although Spanish-origin roots can trigger the nasalisation of prefixes and proclitics in PG, they do so at a significantly lower rate than PG-origin roots.

Nasal harmony and progressive nasalisation are dependent on different mechanisms, consistent with proposals made by Lapierre & Michael (2018) and Estigarribia (2021), and counter to the claim that PG has bidirectional harmony (Lunt 1973; Goldsmith 1976). Regressive nasal harmony in PG can be straightforwardly handled as a synchronically active phenomenon arising through the combination of two different nasalisation processes: agreement of adjacent syllable nuclei and coarticulation within syllables (Thomas 2014; Russell 2022a). I assume that variation in nasal harmony arises from reweighting of phonological constraints, though I leave the specifics of this analysis for future work. I propose that progressive nasalisation, on the other hand, is best accounted for as suppletive allomorphy, in which each suffix or enclitic that productively nasalises is associated with both an oral and a nasal allomorph. The actual forms of the allomorphs are not predictable from the synchronic phonology, and are instead morpheme-specific vestiges of diachronic nasalisation processes. Specifically, root-initial voiceless stop pre-nasalisation is attributable to a historical nasalisation process that has ceased to be productive in PG (Estigarribia 2021; Russell 2023). Rates of progressive nasalisation therefore represent different rates of selection of each allomorph, which are particular to the specific suffix or enclitic.

Findings about the differences between regressive and progressive nasalisation in PG have implications for our understanding of directionality in harmony. The default direction of harmony has been found to be regressive (Hansson 2001, 2010; Hyman 2002), which is reflected in PG. Directionality has also been argued to follow from morphological structure (Baković 2000, 2003). In such a proposal, harmony is stem-controlled, operating from the root outwards to affixes. This, therefore, predicts that prefixing languages will exhibit regressive harmony, suffixing languages will exhibit progressive harmony, and languages with both prefixes and suffixes will exhibit bidirectional harmony. The PG nasalisation system I have described here appears to constitute a counterexample to this last prediction, as PG has both prefixes and suffixes, and yet I argue that the observed patterns are *not* reflections of bidirectional harmony.¹¹ I propose that a more appropriate formulation of the proposal would involve invoking prosodic structure as well as, or in place of, morphological structure. The domain of nasal harmony in PG is the prosodic word, which includes prefixes but excludes suffixes. If directionality follows from prosodic structure, harmony is predicted to apply at the prosodic word level; such an analysis would not predict progressive nasal harmony in PG, which is indeed borne out here. Further research is necessary to assess the typological validity of this prosodic analysis of the directionality of harmony.

Though languages with vowel and/or consonant harmony systems are widespread around the world (Rose & Walker 2011), many have never been examined through quantitative studies of variable harmony application rate, as I have presented here. A close examination of the factors that condition variation in harmony provides invaluable insight into the mechanisms underlying harmony. In the existing literature, several factors have been found to significantly affect the application rate of harmony across typologically diverse languages, including distance between trigger and target in Navajo (Martin 2005; Palakurthy 2021), distance in terms of morphological template in Tommo So (McPherson & Hayes 2016), and root token frequency in Uyghur (Mayer 2005). The present study of PG nasal harmony supports these previous findings, as well as adding morphosyntactic attachment as another relevant factor.

¹¹ As a reviewer points out, though, the source of this apparent counterexample could be attributed to an asymmetry within the morphology of PG itself, as the language can be considered to be more heavily prefixing than suffixing.

The application rate of sibilant harmony in Navajo (nav, Athabaskan, USA) decreases with distance (measured in syllables) between the trigger and target (Martin 2005). The results of this corpus-based study of PG nasalisation parallel this finding, although distance is relevant only in the case of Spanish-origin roots. The importance of distance between the trigger and target of harmony is not particularly surprising, as cases of long-distance harmony are quite rare (Rose & Walker 2011) – and a greater distance between trigger and target could lessen the effects of functional motivations for harmony, such as speech planning and coarticulation. In common analyses of harmony, such as Agreement by Correspondence (Rose & Walker 2004), which treat harmony processes as a form of featural agreement between segments, correspondence relations exist between similar segments, such that corresponding segments agree on the surface with respect to some feature, such as nasality. Correspondence between segments may be limited at a specified distance (Bennett 2015; Shih & Inkelas 2019). In order to model a pattern like that of nasal harmony with Spanish-origin items in PG, distance may thereby be represented, potentially gradiently, as a factor in the application of harmony.

Vowel harmony in suffixes in Tommo So (dto, Dogon, Mali) applies with diminishing frequency in outer layers of the morphology (McPherson & Hayes 2016). Unlike the findings described for Navajo (and here for PG), the Tommo So pattern is not necessarily connected to distance in number of segments or syllables, but rather distance in terms of layers of the morphology. In PG, though, differential application of nasalisation to individual morphemes appears to be most closely connected to the distinction between cliticisation and affixation, rather than to the slot in the verbal template (cf. (2)). The negative prefix, for instance, is further removed from the root than agreement in terms of the verbal template, but nasalises at a higher rate than the Set B agreement morphemes. Additionally, the two sets of agreement morphemes compete for a single slot (Velázquez-Castillo 1991), and yet display significantly different rates of nasalisation. However, if we disregard the exceptional Set B morphemes, PG nasal harmony interactions with prefixes do appear to follow the Tommo So pattern, although nasalisation rates are quite similar across the board. The negation prefix, which is the outermost layer of the morphology examined in this study, has the lowest nasalisation rate, at 91.7%. Additionally, the epenthetic /j/, which is the innermost layer of morphology examined in this study, has the highest nasalisation rate, at 98%.

Backness harmony in Uyghur (uig, Turkic, China) is correlated with the token frequency of the root (Mayer 2005). In Uyghur, a vowel raising process converts harmonic vowels into transparent vowels, rendering the harmony pattern opaque. The rate of opaque harmony for a root is predicted by its token frequency: the more frequent a root is overall, the more likely it is to display opaque harmony. In this study, I have found frequency to be relevant for PG nasal harmony as well: the more frequent the root, the more likely it is to trigger nasal harmony. The interactions between frequency and harmony application rate present an opportunity for interesting future connections to the larger literature regarding frequency effects on phonological processes (e.g. Bybee 2002; Anttila 2006; Coetzee 2008; Coetzee & Kawahara 2012).

The PG corpus data shed light on another factor conditioning variation which has not as yet been discussed in the literature: different types of morphosyntactic attachment reflect distinct harmony patterns. I have shown that Set B morphemes nasalise at a significantly lower rate than all other prefixes. This difference is not attributable to any different phonological property of Set B morphemes, and instead may reflect a morphological difference between prefixation and cliticisation. Questions still remain, however, regarding the behaviour of specific morphemes as well as the interactions between root origin and Set B morphemes. The 2SG.B proclitic, for instance, does nasalise at a lower rate than prefixes, but to a lesser extent: it is possible that frequency effects could be in play here, as tokens of the 2SG.B proclitic are quite frequent in the corpus, particularly compared to other Set B proclitics. Further investigation of the interactions of PG morphosyntax and phonology is necessary; regardless, this data make an important contribution to the literature in demonstrating that morphosyntax could be a significant factor conditioning variation of harmony application.

Finally, these findings run counter to several persistent language ideologies. None of the social factors included in this study – gender, age and community affiliation (rural or urban) – were found to significantly improve model fit for any of the models presented here. Within Paraguay, ideologies of linguistic purism abound, in which rural communities are associated with ‘true’ PG, and urban communities with bilingualism and language mixing with Spanish (Rubin 1968; Gómez-Rendón 2007). Additionally, rural communities in Paraguay are reported to have a higher degree of language loyalty to Guaraní than urban communities (Solé 1991; Gynan 1998). Ideologies about the relative purity of rural vs. urban PG go hand in hand with attitudes related to age. As urban areas are associated with youth in Paraguay, there is an expectation that younger generations of people in Paraguay are more likely to be bilingual in PG and Spanish, and more likely to be exposed to bilingual speech (Bittar Prieto 2021). Given that nasal harmony is a characteristic property of PG, and is absent in Spanish, one might hypothesise that rates of nasalisation would be higher for speakers in rural communities (and for older speakers) than for young speakers and those in urban communities. However, again, no social factors were found to significantly contribute to predicted nasalisation rate at all.¹² These findings are informative in that they potentially counter widespread ideologies, deconstructing the artificial construction of an ‘idyllic rural space’ and veneration of the speech of elders as more pure (Gordon 2019). It remains to be seen whether this lack of differentiation according to age and setting holds for other phonetic, phonological and morphosyntactic factors in PG.

7. Conclusion

Investigating a corpus of sociolinguistic interviews sheds light on the complex nature of nasalisation in PG. The two types of nasalisation in PG – regressive and progressive – are in fact distinct, and actual rates of progressive and regressive nasalisation significantly differ from each other. I argue that regressive nasal harmony is synchronically active, while morpheme-specific progressive nasalisation is phonologically conditioned suppletive allomorphy. Several different factors contribute to the observed nasalisation rate, including the direction of nasalisation, target morpheme identity, root etymological origin, root token frequency, relationship to stress and distance between the trigger and the target. The etymological origin of the root (Spanish vs. PG) significantly affects the rate of progressive, but not regressive nasalisation, signalling that harmony has been extended to Spanish-origin items. Various phonological factors, like stress and distance, are relevant in determining the rate of nasalisation triggered by a Spanish-origin root. Regressive nasal harmony additionally interacts with the morphosyntax: Set B agreement morphemes, which have been analysed as clitics, nasalise at a significantly lower rate than all other prefixes. These findings contribute to a deeper understanding of nasalisation within PG and of the factors that condition variable harmony application rate across languages.

¹²As pointed out by a reviewer, the relevance of social factors could potentially have been obscured by a difference in the number of Guaraní- versus Spanish-origin roots used by rural speakers as opposed to urban ones: for instance, rural speakers might have favoured more Guaraní-origin roots overall. However, in the corpus consulted for this study, rural speakers used 976 Guaraní-origin roots (51.5%) and 919 Spanish-origin roots (48.5%), while urban speakers used 988 Guaraní-origin roots (56.5%), and 758 Spanish-origin roots (43.4%).

A. Abbreviations

1	first person
2	second person
3	third person
A	Set A agreement prefix
B	Set B agreement prefix
AGD	agent demoter
ATTEN	attenuative
CAUS	causative
COLL	collective
COND	conditional
DESID	desiderative
DIM	diminutive
DOM	differential object marking
EXCL	exclusive
FRUS	frustrative
FUT	future
H	relational /h/
INCIP	incipient
INCL	inclusive
INTENS	intensifier
J	epenthetic /j/
LOC	locative
NEG	negative
NMLZ	nominaliser
N.FUT	nominal future
N.PST	nominal past
PASS	passive
PL	plural
Q	question
R	relational /r/
RECIP	reciprocal
REL	relativiser
SG	singular
TOTAL	totalitative
TRANS	transitiviser

B. Figures

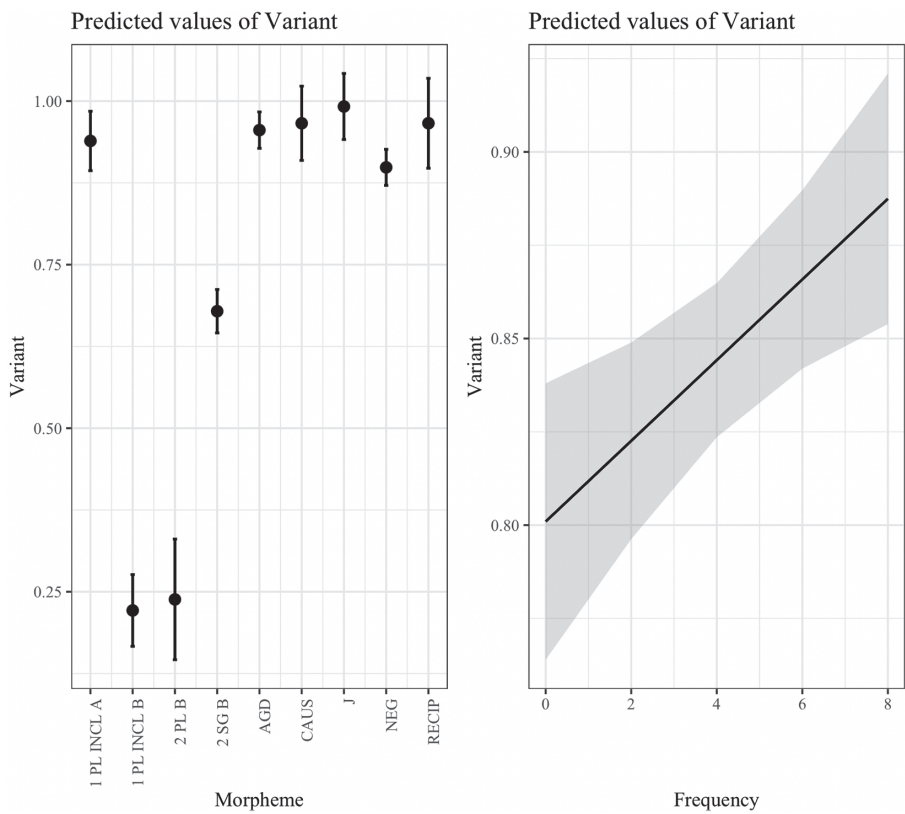


Figure B1. Predicted probabilities of the nasal variant for significant main effects in regressive nasal harmony (see Table 12).

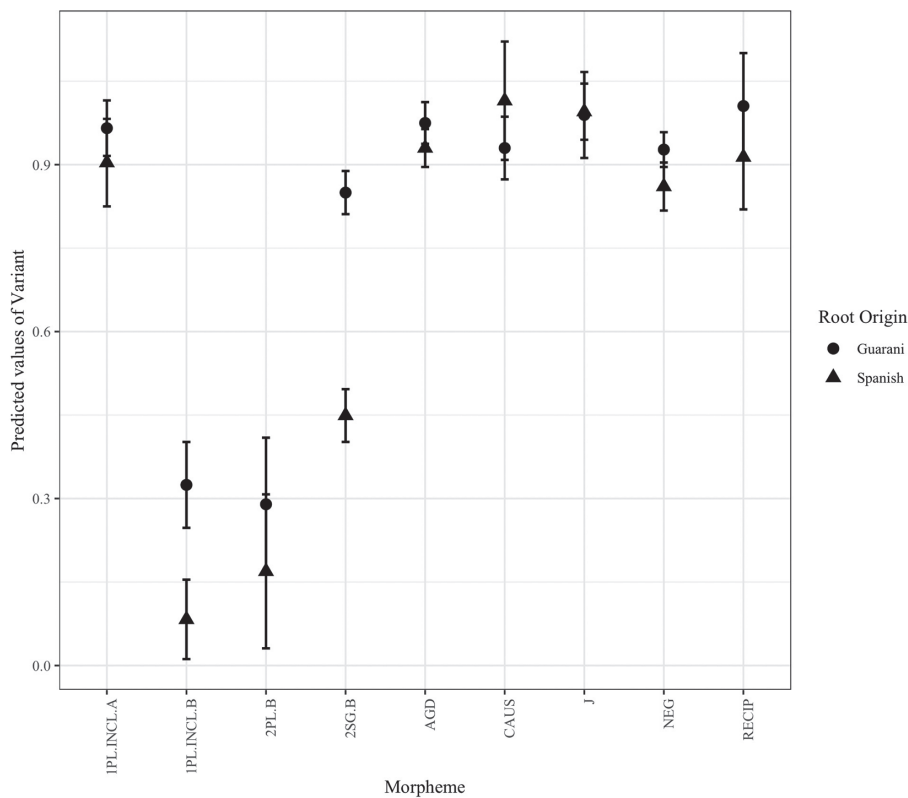


Figure B2. Predicted probability of nasal variant for the interactions of root etymological origin and target morpheme identity in regressive nasal harmony (see Table 12).

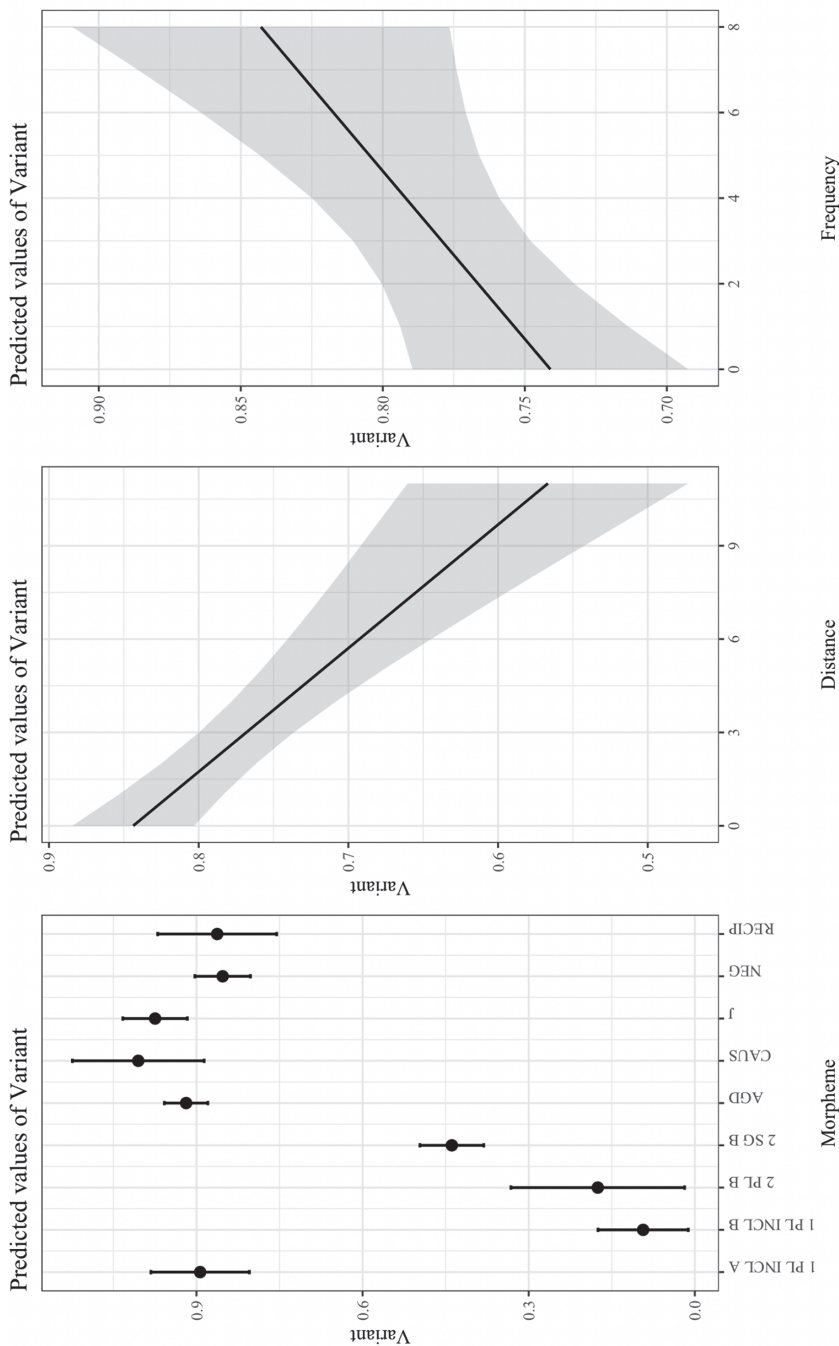


Figure B3. Predicted probability of nasal variant for the significant main effects in regressive nasalisation triggered by Spanish-origin roots (see Table 14).

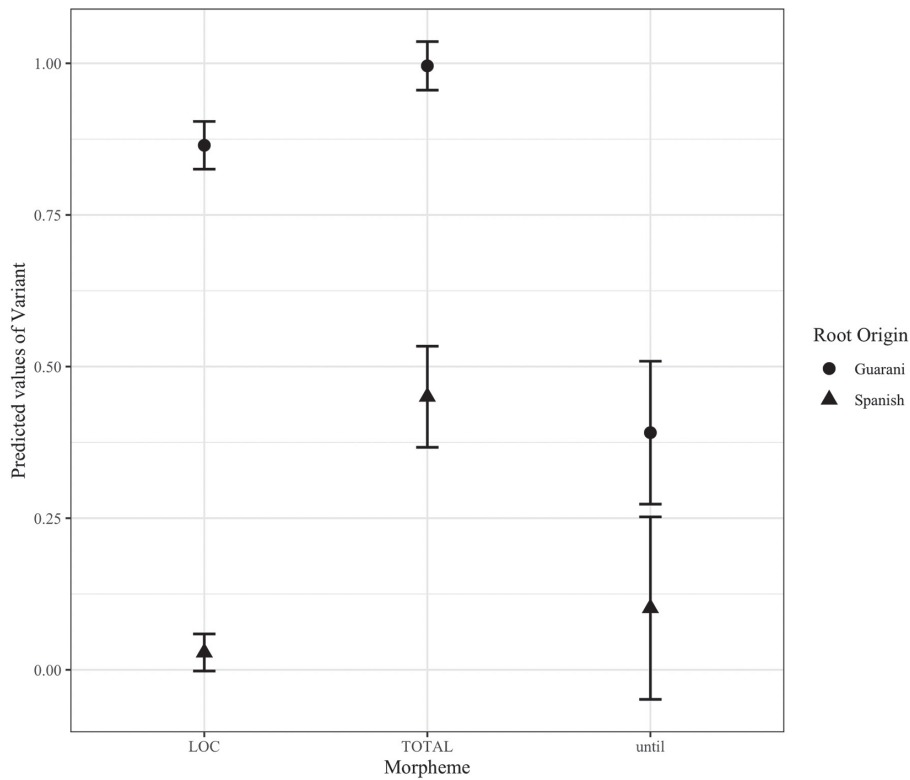


Figure B4. Predicted probability of nasal variant for the interactions of root etymological origin and target morpheme identity in progressive nasalisation (see Table 15).

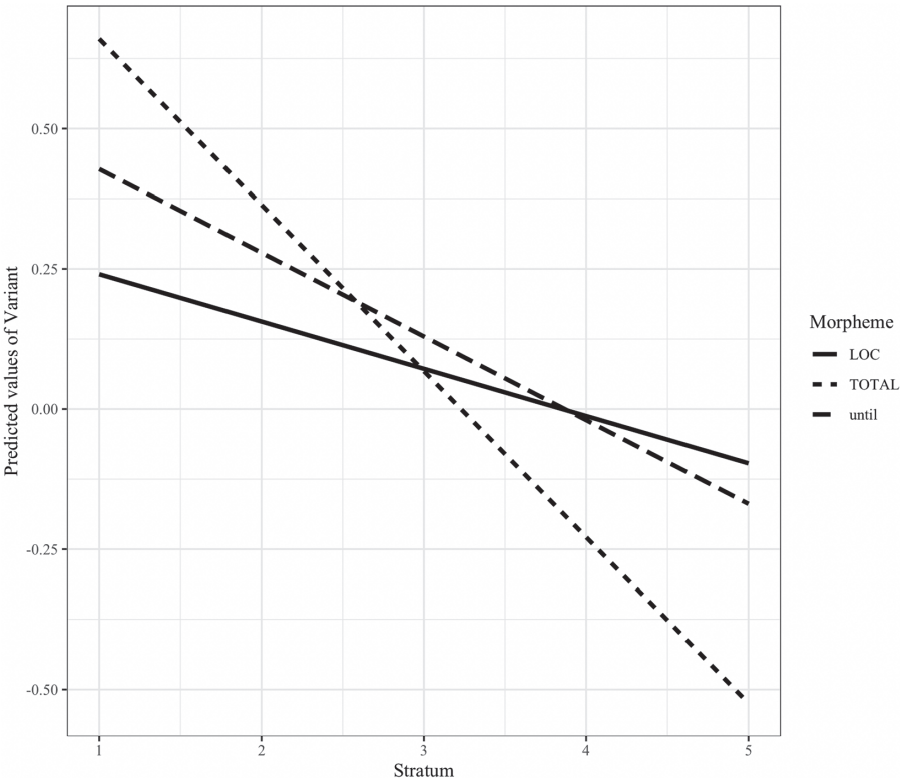


Figure B5. Predicted probability of nasal variant for the interactions of root lexical stratum and target morpheme identity in progressive nasalisation triggered by Spanish-origin roots (see Table 17).

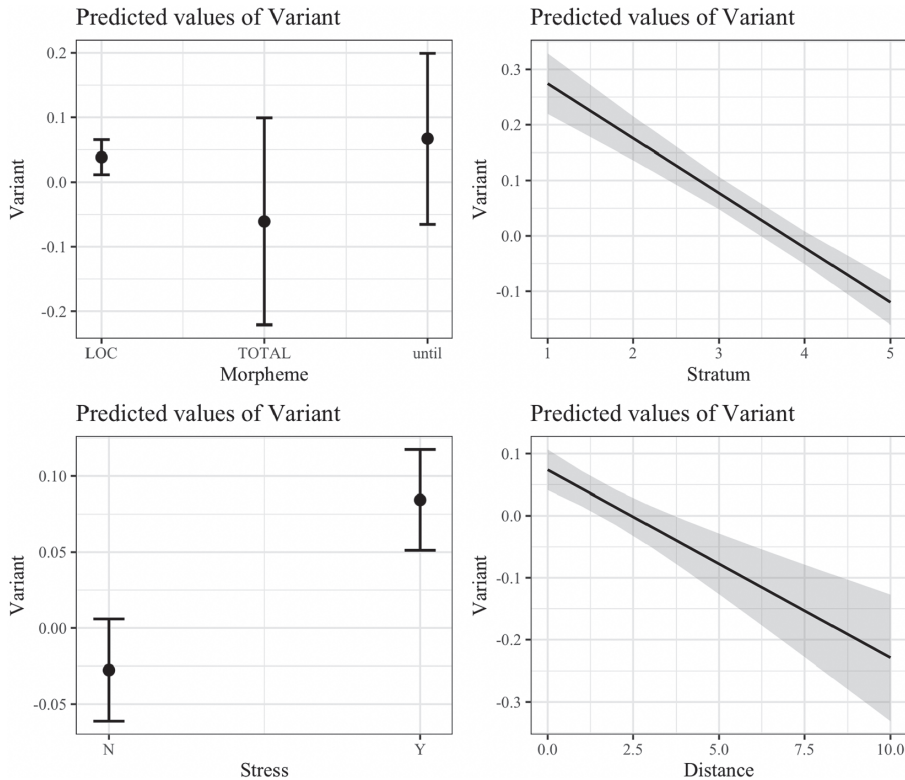


Figure B6. Predicted probability of nasal variant for the significant effects in progressive nasalisation triggered by Spanish-origin roots (see Table 17).

Acknowledgements. I am immensely grateful to Maria Gómez and Irma Ovelar Easty, who have been so gracious in sharing so much of their time and knowledge with me – *aguyjeve!* A huge thank you to Josefina Bittar for her generosity in sharing this incredible resource of Guaraní data with me. Thanks to Lev Michael, Isaac Bleaman, Hannah Sande and Maksymilian Dąbkowski, as well as audiences at the LSA and three anonymous reviewers for comments and suggestions on various aspects of this work. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1752814. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Competing interests. The author declares none.

References

- Adelaar, Willem F. H. (1994). The nasal/oral distinction in Paraguayan Guaraní. *Revista Latinoamericana de Estudios Etnolingüísticos* 8, 125–133.
- Anttila, Arto (2006). Variation and opacity. *NLLT* 24, 893–944.
- Ayala, Valentín (1996). *Gramática Guaraní*. Buenos Aires: Edición del Ministerio de Cultura y Educación de la República Argentina.
- Bakker, Dik, Jorge A. Gómez Rendón & Ewald Hekking (2008). Spanish meets Guaraní, Otomí, and Quechua: a multilingual confrontation. In Thomas Stolz, Dik Bakker & Rosa Salas Palomo (eds.) *Aspects of language contact: new theoretical, methodological and empirical findings with special focus on Romancisation processes*. Berlin: Mouton de Gruyter, 165–238.
- Baković, Eric (2000). *Harmony, dominance and control*. PhD dissertation, Rutgers University.
- Baković, Eric (2003). Vowel harmony and stem identity. *San Diego Linguistics Papers* 1, 1–42.
- Baraúna, Fabíola Azevedo (2020). *Tendência tipológica de harmonia nasais e palatização em línguas indígenas brasileiras*. PhD dissertation, Universidade Federal do Pará, Belém.
- Barratt, Leslie (1981). Prenasalised stops in Guaraní: where the autosegment fails. *Linguistic Analysis* 7, 187–202.

- Bartoń, Kamil (2022). *Mu-MIn: multi-model inference*. R package. Available at <https://CRAN.R-project.org/package=MumIn>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.
- Beckman, Jill (1998). *Positional faithfulness*. PhD dissertation, University of Massachusetts Amherst.
- Bennett, William (2015). Assimilation, dissimilation, and surface correspondence in Sundanese. *NLLT* **33**, 371–415.
- Bittar Prieto, Josefina (2021). *Spanish-origin verbal and grammatical borrowing in Paraguayan Guaraní*. PhD dissertation, University of New Mexico.
- Boidin, Capucine (2006). Jopara: una vertiente sol y sombra del mestizaje. In Wolf Dietrich & Haralambos Symeonidis (eds.) *Guaraní y “Maweti-Tupi-Guarani”: estudios históricos y descriptivos sobre una familia lingüística de América del Sur*. Munster: LIT-Verlag, 303–331.
- Bybee, Joan L. (2002). Word frequency and context of use in lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* **14**, 261–290.
- Cabral, Ana Suely Arruda Câmara & Aryon Dall’Igna Rodrigues (2011). The interface of stress and nasality in Tupi-Guarani languages in a historical perspective. *Rivista di Linguistica* **7**, 1–15.
- Cardona, Maria Zulma (2008). *The convergence, emergence and transcendence of Jopara: a third language in bilingual Paraguay*. Master’s thesis, California State University San Marcos.
- Cardoso, Valéria Faria (2009). Sistematização da fonologia kaiojá: nazalização e/ou oralização. *Sínteses* **14**, 31–72.
- Cassano, Paul (1971). The attribution of vocalic nasalisation in Paraguayan Spanish to Guaraní influence. *Romance Notes* **13**, 190–192.
- Choi, Jinny K. (2003). Language attitudes and the future of bilingualism: the case of Paraguay. *International Journal of Bilingual Education and Bilingualism* **6**, 81–94.
- Choi, Jinny K. (2004). Language planning and language spread of Guaraní in Paraguay: comparison, evaluation, and implication. *Language Problems and Language Planning* **28**, 241–259.
- Choi, Jinny K. (2005). The linguistic situation in urban Paraguay: a tendency toward Spanish–Guaraní bilingualism? *Spanish Context* **2**, 175–201.
- Clements, G. N. & Engin Sezer (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*, volume II. Dordrecht: Foris, 213–255.
- Coetzee, Andries (2008). Phonological variation and lexical frequency. *NELS* **38**, 1–28.
- Coetzee, Andries & Shigeto Kawahara (2012). Frequency biases in phonological variation. *NLLT* **31**, 47–89.
- Daviet, Windy (2016). *Observations sociolinguistiques et analyse de la phonologie du dialecte ava du guaraní bolivien, langue tupi-guaraní de Bolivie*. Master’s thesis, Université Lumière Lyon 2.
- DGEEC, Dirección General de Estadística, Encuestas y Censos (2012). Censo nacional de población y viviendas 2012 [National population and housing census of 2012]. Available at <https://www.ine.gov.py/default.php?publicacion=33>.
- Dietrich, Wolf (2010). Lexical evidence for a redefinition of Paraguayan ‘Jopara’. *Language Typology Universals* **63**, 39–51.
- Dabkowski, Maksymilian (2022). Prosody drives Paraguayan Guaraní suffix order. In Peter Jurgec, Liisa Duncan, Emily Elfner, Yoonjung Kang, Alexei Kochetov, Brittney K. O’Neill, Avery Ozburn, Keren Rice, Nathan Sanders, Jessamyn Schertz, Nate Shaftoe & Lisa Sullivan (eds.) *Supplemental proceedings of the 2021 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America, 12 pp.
- Estigarribia, Bruno (2015). Guaraní–Spanish Jopara mixing in a Paraguayan novel. *Journal of Language Contact* **8**, 183–222.
- Estigarribia, Bruno (2020). *A grammar of Paraguayan Guaraní*. London: University College London Press.
- Estigarribia, Bruno (2021). A diachronic account of exceptional progressive nasalisation patterns in Guaraní causatives. *IJAL* **87**, 203–241.
- Fernández Barrera, Meritxell (2015). *Paraguayan Guaraní: some considerations about language mixing and an acoustic study of urban and rural vowels*. Master’s thesis, Leiden University.
- Goldsmith, John A. (1976). *Autosegmental phonology*. PhD dissertation, Massachusetts Institute of Technology.
- Gómez, María, Irma Easty Ovelar, Madeline Bossi, Maksymilian Dabkowski, Emily Drummond, Emily Grabowski, Rebecca Jarvis, Phuong Khuu, Lev Michael & Katherine Russell (2020). Berkeley field methods: Paraguayan Guaraní. In *California language archive*. Berkeley, CA: University of California, Berkeley. Available at <https://doi.org/10.7297/X2PR7TNF>.
- Gómez-Rendón, Jorge (2007). Grammatical borrowing in Paraguayan Guaraní. In Yaron Matras & Jeanette Sakel (eds.) *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter Mouton, 523–550.
- Gordon, Matthew J. (2019). Language variation and change in rural communities. *Annual Review of Linguistics* **5**, 435–453.
- Gregores, Emma & Jorge A. Suárez (1967). *A description of colloquial Guaraní*. The Hague: Mouton.
- Gynan, Shaw N. (1998). Attitudinal dimensions of Guaraní–Spanish bilingualism in Paraguay. *Modern & Classical Languages* **61**, 35–59.
- Gynan, Shaw N. (2011). Linguistic demography and attitudinal dimensions of intergenerational transmission of Guaraní and Spanish in Paraguay. *Southwest Journal of Linguistics* **30**, 57–94.
- Hamidzadeh, Khashayar & Kevin Russell (2014). The phonology of reduplication in Paraguayan Guaraní. In Natalie Weber & Sihwei Chen (eds.) *Proceedings of the Workshop on Structure and Constituency in the Languages of the Americas 19*. Vancouver, BC: UBC Working Papers in Linguistics, 95–108.

- Hansson, Gunnar (2001). *Theoretical and typological issues in consonant harmony*. PhD dissertation, University of California, Berkeley.
- Hansson, Gunnar (2010). *Consonant harmony: long-distance interactions in phonology*. Berkeley, CA: University of California Press.
- Humbert, Helga & Glyne Piggott (1997). Representations and constraints: the case of Guaraní nasal harmony. In Geert Booij & Jeroen van de Weijer (eds.) *HIL phonology papers III*. The Hague: Holland Academic Graphics, 219–256.
- Hyman, Larry (2002). *Is there a right-to-left bias in vowel harmony?* Ms., University of California, Berkeley.
- Itô, Junko & Armin Mester (1999). The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*. Oxford: Blackwell, 62–100.
- Jensen, Cheryl (1998). Comparative Tupi-Guaraní morphosyntax. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *The handbook of Amazonian languages*. Berlin: Mouton de Gruyter, 490–618.
- Kaiser, Eden (2008). Nasal spreading in Paraguayan Guaraní: introducing long-distance continuous spreading. *Amerindia* 32, 283–300.
- Kertész, Zsuzsa (2003). Vowel harmony and the stratified lexicon of Hungarian. *The Odd Yearbook* 7, 62–77.
- Lapierre, Myriam & Lev Michael (2018). Nasal harmony in Tupí-Guaraní: a comparative synthesis. Paper presented at SSLA III, Amherst, MA, 19–21 October 2018.
- Lunt, Horace G. (1973). Remarks on nasality: the case of Guaraní. In Stephen R. Anderson (ed.) *A Festschrift for Morris Halle*. Austin, TX: Holt, Rinehart and Winston, 131–139.
- Lustig, Wolf (2010). Mba'êchapa oiko la guarani? Guaraní y jopara en el Paraguay. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico* 4, 19–43.
- Martin, Andrew T. (2005). *The effects of distance on lexical bias: sibilant harmony in Navajo compounds*. Master's thesis, University of California, Los Angeles.
- Mayer, Connor (2005). *Issues in Uyghur backness harmony: corpus, experimental, and computational studies*. PhD dissertation, University of California, Los Angeles.
- McPherson, Laura & Bruce Hayes (2016). Relating application frequency to morphological structure: the case of Tommo So vowel harmony. *Phonology* 33, 125–167.
- Melià, Bartomeu (1974). Hacia una 'tercera lengua' en el Paraguay. *Estudios Paraguayos* 2, 31–71.
- Miranda, Camille Cardoso (2018). *Estudo comparativo do fenômeno de nasalização em línguas da família Tupi-Guaraní (tronco Tupi)*. PhD dissertation, Universidade Federal do Pará.
- Miranda, Camille Cardoso & Gessiane Lobato Picanço (2020). O fenômeno de nasalização em línguas Tupi-Guaraní. *LIAMES: Línguas Indígenas Americanas* 20, 1–21.
- Mistieri, Fernanda Regina (2013). *Acento em línguas tupi-guarani: uma análise comparada*. Master's thesis, Universidade Estadual Paulista, Araraquara.
- Morinigo, Marcos A. (1959). Influencia del español en la estructura lingüística del guaraní. *Filología* 5, 235–247.
- Nichols, Johanna (1992). *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.
- Palacios Alcaine, Azucena (2008). Paraguay. In Azucena Palacios Alcaine (ed.) *El Español en América: contactos lingüísticos en Hispanoamérica*. Barcelona: Ariel, 279–300.
- Palakurthy, Kayla (2021). The status of sibilant harmony in Diné Bizaad (Navajo). *JIPA* 53, 404–422.
- Piggott, Glyne (1992). Variability in feature dependency: the case of nasality. *NLLT* 10, 33–78.
- Pinta, Justin (2013). *Lexical strata in loanword phonology: Spanish loans in Guaraní*. Master's thesis, University of North Carolina at Chapel Hill.
- Pinta, Justin & Jennifer L. Smith (2017). Spanish loans and evidence for stratification in the Guaraní lexicon. In Bruno Estigarribia & Justin Pinta (eds.) *Guaraní linguistics in the 21st century*. Leiden: Brill, 285–314.
- Puthaval, Sarala (2013). *Loanwords, prominence and the basis for Mongolian vowel harmony*. Master's thesis, University of Washington.
- Ringen, Catherine & Orvokki Heinämäki (1999). Variation in Finnish vowel harmony: an OT account. *NLLT* 17, 303–337.
- Romaine, Suzanne (1995). *Bilingualism*. Oxford: Blackwell.
- Rose, Françoise (2008). *Grammaire de emérillon teko, une langue tupi-guarani de Guyane française*. Leuven: Peeters.
- Rose, Sharon & Rachel Walker (2004). A typology of consonant agreement as correspondence. *Lg* 80, 475–531.
- Rose, Sharon & Rachel Walker (2011). Harmony systems. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*, 2nd edition. Oxford: Blackwell, 240–290.
- Rubin, Joan (1968). *National bilingualism in Paraguay*. The Hague: Mouton.
- Russell, Katherine R. (2022a). Interactions of nasal harmony and word-internal language mixing in Paraguayan Guaraní. *Languages* 7, 1–24.
- Russell, Katherine R. (2022b). Reduplication in Paraguayan Guaraní: a novel prosodic analysis. In Noah Elkins, Bruce Hayes, Jinyoung Jo & Jian-Leat Siah (eds.) *Supplemental proceedings of the 2022 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America, 13 pp.
- Russell, Katherine R. (2023). Nasalization in Paraguayan Guaraní. In Marianne Huijsmans, D. K. E. Reisinger & Rose Underhill (eds.) *Proceedings of the Workshop on Structure and Constituency in the Languages of the Americas 25*. Vancouver, BC: UBC Working Papers in Linguistics, 131–145.

- Shih, Stephanie & Sharon Inkelas (2019). Autosegmental aims in surface-optimizing phonology. *LI* **50**, 137–196.
- Siewierska, Anna (1996). Word order type and alignment type. *Sprachtypologie und Universalienforschung* **49**, 149–176.
- Solé, Maria-Josep (1992). Phonetic and phonological processes: the case of nasalization. *Language Speech* **35**, 29–43.
- Solé, Yolanda R. (1991). The Guaraní–Spanish situation. *The Georgetown Journal of Languages and Linguistics* **2**, 297–348.
- Stanton, Juliet (2017). *Constraints on the distribution of nasal–stop sequences: an argument for contrast*. PhD dissertation, Massachusetts Institute of Technology.
- Steriade, Donca (1993). Closure, release and nasal contours. In Marie K. Huffman & Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, number 5 in Phonetics and Phonology. San Diego, CA: Academic Press, 401–470.
- Thomas, Guillaume (2014). A split analysis of nasal harmony in Mbyá. *Rivista di Linguistica* **10**, 75–104.
- Thun, Harald (2005). ‘Code switching’, ‘code mixing’, ‘reproduction traditionnelle’ et phénomènes apparentés dans le guaraní paraguayen et dans le castillan du Paraguay. *Rivista di Linguistica* **17**, 311–346.
- Tonhauser, Judith (2007). Nominal tense? the meaning of Guaraní nominal temporal markers. *Lg* **83**, 831–869.
- Trudgill, Peter (1995). *Sociolinguistics: an introduction to language and society*. New York: Penguin.
- Velázquez-Castillo, Maura (1991). The semantics of Guaraní agreement markers. *BLS* **17**, 324–335.
- Walker, Rachel (1999). Guaraní voiceless stops in oral versus nasal contexts: an acoustical study. *JIPA* **29**, 63–94.
- Walker, Rachel (2000). Long-distance consonantal identity effects. Paper presented at the West Coast Conference on Formal Linguistics 19, Los Angeles, CA, 4–6 February 2000.
- Woolford, Ellen (2016). Two types of portmanteau agreement: syntactic and morphological. In Géraldine Legendre, Michael T. Putnam, Henriette de Swart & Erin Zaroukian (eds.) *Optimality-Theoretic syntax, semantics and pragmatics: from uni- to bidirectional optimization*. Oxford: Oxford University Press, 111–135.
- Zajícová, Lenka (2009). *El bilingüismo paraguayo: usos y actitudes hacia el guaraní y el castellano*. Madrid: Vervuert.
- Zubizarreta, María Luisa (2022). The verbal phrase in Paraguayan Guaraní: a case study on the role of prosody in linearization. *Languages* **7**, 1–30.
- Zubizarreta, María Luisa & Roumyana Pancheva (2017). A formal characterization of person-based alignment: the case of Paraguayan Guaraní. *NLLT* **35**, 1161–1204.