DOI: 10.26346/1120-2726-205

A corpus-based approach to map target vowel asymmetry in Brazilian Veneto metaphony

Guilherme D. Garcia,[a] Natália Brambatti Guzzo[b]

[a] Université Laval, Québec City, Canada <guilherme.garcia@lli.ulaval.ca>
[b] Saint Mary's University, Halifax, Canada <nataliaguzzo@me.com>

Metaphony targeting high-mid vowels /e, o/ is a characteristic of Central Veneto, a dialect of Veneto spoken in northeastern Italy. In a closely related understudied dialect spoken in southern Brazil, namely Brazilian Veneto (locally known as Talian), metaphony is also observed. Although the phenomenon is reported as variable for both dialects, little is known about how such variation is structured. In this paper, we explore the structural conditioning of metaphony in Talian through a corpus study. First, we introduce the Talian Corpus, a corpus of written materials in Talian that promotes the linguistic study of this variety. We then show that metaphony in this dialect is asymmetrical, as /e, o/ exhibit different rates of application, which are conditioned by number of syllables in the word and morphology. Finally, we formalize this asymmetry using a MaxEnt Grammar.

KEYWORDS: Brazilian Veneto, Talian, metaphony, corpus data.

1. Introduction

Many Italo-Romance languages exhibit metaphony, which may target mid vowels (high-mid and/or low-mid) as well as low vowels (e.g. Maiden 1987; Savoia & Maiden 1997). In this paper, we discuss the patterns of metaphony observed in Brazilian Veneto (locally known as Talian), an understudied dialect of Veneto spoken in several parts of Brazil that is closely related to Central Veneto. In Talian, like in Central Veneto, metaphony targets stressed high-mid vowels (/e, o/) and is triggered by a posttonic /i/ (in final position in the case of Talian; e.g. /ˈpesi/ → [ˈpisi] 'fish.PL', /ˈkonti/ → [ˈkunti] 'buck.PL (money)'). Like in Central Veneto, metaphony applies variably in Talian (Guzzo 2022), although its conditioning factors are unclear.

Our objective is to explore the structural conditioning of metaphony in Talian through a corpus study based on written data. The data in question were extracted from the Talian Corpus, which includes sentences and words in written Talian obtained from various sources. Since Talian has no standardized orthography (see e.g. Luzzatto 2000), the writers may be following their own intuitions about language use in

their texts. In the case of metaphony (and other variable phenomena), the writers' spelling may be reflecting their own productions. Given that Talian's phoneme inventory is relatively simple (with seven vowels in stressed position; Guzzo 2022), sound-letter correspondence may in effect be used as a tool for representing variable processes in orthography (with, for example, orthographic *i* and *u* being used to indicate metaphony in items such as *pissi* 'fish.PL' and *cunti* 'buck.PL', respectively). Indeed, orthography has been shown to reflect phonological patterns even when spelling is standardized (e.g. Eisenstein 2015).

In what follows, we first discuss the Veneto dialect under study, since its development in Brazil has a number of particularities. In the subsequent sections, we describe metaphony in Talian as well as the tools that we used for developing our corpus and for extracting the metaphony data. Afterwards, we present and discuss our results. As will be shown below, we found an asymmetry in the application of metaphony for target /e/ and /o/, which is conditioned by both phonological and morphological factors. These results are formalized using a MaxEnt Grammar (Goldwater & Johnson 2003; Hayes & Wilson 2008). We conclude by laying out some directions for future research.

2. The Talian language

Talian developed in Brazil with Italian immigration, which started in the mid 19th century. In southern Brazil, which is where most Talian-speaking communities are located, the first settlements of Italian immigrants were founded in 1875. Immigrants initially settled mostly in the region known as Italian Immigration Area (or IIA, which roughly corresponds to the cluster of dots in Figure 1; De Boni & Costa 1979; Frosi & Mioranza 2009), in the state of Rio Grande do Sul. The other southern Brazilian states (namely, Santa Catarina and Paraná), as well as the southeastern states (especially São Paulo and Espírito Santo), also received massive waves of Italian immigrants (see e.g. Loriato 2019).

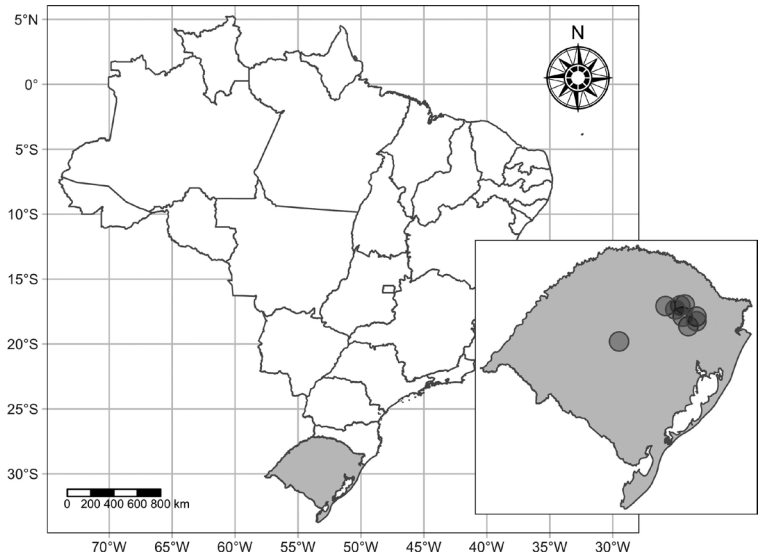

Figure 1. Municipalities where Talian is an official language in the state of Rio Grande do Sul, in southern Brazil.

Most immigrants who settled in the IIA (and also in other areas of Brazil) were from the Veneto region in today's Italy, and spoke a Veneto dialect (see Table 1). Since immigration to southern Brazil was implemented mainly with the objective of occupying scarcely populated areas, most Italian immigrant communities did not have any sustained contact with Portuguese-speaking communities, nor with other immigrant communities that had settled in neighboring territories (De Boni & Costa 1979; Frosi & Mioranza 2009). In addition, land assignment followed immigrants' order of arrival, which resulted in many Italian communities exhibiting a mixture of languages and dialects (Frosi & Mioranza 1983).

REGION OF ORIGIN	%
Veneto	54
Lombardia	33
Trentino–Alto Adige	7
Friuli–Venezia Giulia	4.5
Others	1.5

Table 1. Regions of origin of Italian immigrants (Frosi & Mioranza 2009).

These factors promoted the development of a Veneto-based koine in these communities (Frosi & Mioranza 1983).[1] This koine (Talian) shares many similarities with other Veneto dialects, especially Central Veneto (see e.g. Guzzo 2022; Frasson 2020; this dialect is also referred to as Padovano-Vicentino-Polesano, Zamboni 1974).[2] Regarding its phonology, Talian is similar to other Veneto dialects in that it has a seven-vowel system in stressed position (/i, e, ɛ, u, o, ɔ, a/), a five-vowel system in pretonic position (/i, e, u, o, a/), and a trisyllabic window for stress assignment. Talian also shares most of its consonants with other Veneto varieties (except for bilabial and dental fricatives, which are present in a few varieties but absent in Talian; Guzzo 2022; see also Zamboni 1974). Like Central Veneto, Talian exhibits metaphony of stressed /e, o/. In the next section, we describe the process of metaphony in Talian, tracing parallels with its application in Central Veneto.

3. Metaphony in Talian

In many Romance languages, metaphony may target various vowel heights, from low to high-mid (see e.g. Maiden 1987; Savoia & Maiden 1997). In the Veneto varieties that exhibit metaphony, namely Central Veneto and Talian, the phenomenon is much more constrained, in that it is limited to targeting stressed /e, o/, and it is triggered by posttonic /i/ (Zamboni 1974, Frosi & Mioranza 1983, Belloni 2009, Walker 2005, 2010, Perrone 2016, Guzzo 2022). In Talian, metaphony seems to be triggered only by final /i/, while in Central Veneto the trigger may also be in non-final position (Walker 2005).

In both Talian and Central Veneto, the trigger vowel is usually a separate morpheme – either the plural masculine morpheme or the second person singular inflection, as exemplified in (1). However, metaphony may also apply in monomorphemic words (e.g. /ˈd͡ʒeɾi/ → [ˈd͡ʒiɾi] ‘yesterday’, /ˈloɾi/ → [ˈluɾi] ‘they.MASC’). The examples in (1) are from Talian – the target stressed vowels are in bold, while unstressed vowels that can also be targeted by the process (i.e. in words with antepenultimate stress) are underlined. All final /i/s in the words in (1) and in the examples that follow correspond to a separate morpheme.

(1) a. Antepenultimate stress
ˈz**o**ve̲ni ~ ˈz**u**vi̲ni ‘young.PL’
ˈgɾ**o**sto̲li ~ ˈgɾ**u**stu̲li ‘type of pastry.PL’
b. Penultimate stress
ˈ**o**vi ~ ˈ**u**vi ‘egg.PL’
ˈko**ɾ**i ~ ˈku**ɾ**i ‘run.2SG’

	ˈ**pe**si ~ ˈ**pi**si	‘fish.PL’
	ˈ**be**vi ~ ˈ**bi**vi	‘drink.2SG’
c.	Final stress	
	faˈz**oi** ~ faˈz**ui**	‘bean.PL’ (SG /fazol/)
	niˈs**oi** ~ niˈs**ui**	‘bedsheet.PL’ (SG /nisol/)

As the examples in (1) suggest, metaphony is variable in Talian, similarly to what is observed in Central Veneto (Walker 2005). In addition, the phenomenon applies to any stress position within the trisyllabic window, as well as to posttonic syllables with /e, o/. In the case of words with antepenultimate stress, metaphony may apply to the unstressed non-final syllable when there is no stressed high-mid vowel, as exemplified in (2). However, non-high-mid stressed vowels are not affected.

(2)	ˈɔm**e̱**ni ~ ˈɔm**i̱**ni, and not *ˈom**i̱**ni, *ˈum**i̱**ni	‘man.PL’
	ˈpɛrs**e̱**gi ~ ˈpɛrs**i**gi, and not *ˈpers**i**gi, *ˈpirs**i**gi	‘peach.PL’

Regarding metaphony in final syllables, in Talian the process seems to display a target vowel asymmetry, since it appears to apply only with stressed /o/, but not with stressed /e/. In other words, in examples with target /e/ equivalent to those in (1c), metaphony does not apply (e.g. /kaˈvei/ → *[kaˈvii] ‘hair.PL’, from /kaˈvel/ ‘hair.SG’). This seems to contrast with Central Veneto, where forms such as [ˈkɾii] ‘believe.2SG’ (alternating with [ˈkɾei]) are found (Walker 2005).[3]

Metaphony may also variably spread to pretonic vowels. This has been observed in both Talian and Central Veneto (Walker 2005, Guzzo 2022), as illustrated in (3). The examples in (3) are also from Talian. With pretonic vowels, a target vowel asymmetry also seems to be found, given that pretonic /o/ tends to undergo metaphony more frequently than pretonic /e/ (Walker 2005).

(3)	boˈtoni ~ boˈtuni ~ buˈtuni	‘button.PL’
	omeˈneti ~ omeˈniti ~ omiˈniti	‘man.DIM.PL’

Regarding the examples in (3), two additional observations should be made. The first one is that metaphony applies to both root vowels and suffix vowels (e.g. in [omeˈneti] and its related forms, the stressed vowel is in the diminutive suffix, whose singular form is [-eto]). The second observation is that, in words with multiple pretonic vowels (such as [omeˈneti]), metaphony spreading to pretonic position may not target all pretonic vowels.

As indicated above, metaphony in Talian (as well as in Central Veneto) is a variable process. Although the main conditioning factors

for the application of metaphony are well understood (i.e. stress, quality of the target and trigger vowels), little is known about how the variation involving metaphony is structured in the language. It is uncertain whether the variation is constrained by segmental context (such as the type of consonant that precedes or follows the target vowel), prosodic factors (such as number of syllables in the word), or morphological factors (such as the target vowel being in the root or in a suffix). In addition, it is unclear whether the target vowel asymmetries observed with metaphony in final and pretonic syllables are also found in other positions.

As previously mentioned, in this paper, we probe the factors that condition metaphony in Talian by examining data from a corpus of written materials (the Talian Corpus). Before we proceed to the discussion of the patterns found in our data, we describe how the corpus was constituted, as well as how the relevant data were extracted from it.

4. The Talian Corpus

The first challenge in creating a corpus of written materials for any understudied language is to gather analyzable data, which will often require scanning and digitizing physical texts. Our texts consist of book excerpts as well as newspaper articles published in *Correio Riograndense*, one of the oldest newspapers in Brazil (founded in 1909, with headquarters in Caxias do Sul, in the state of Rio Grande do Sul), and in *O Florense* (founded in 1986, with headquarters in Flores da Cunha, Rio Grande do Sul).[4] The language of these newspapers is Portuguese; however, both have sections (usually of one page or column) written in Talian. The articles from *O Florense* are written by a single author, whereas the articles from *Correio Riograndense* and the book excerpts are authored by various individuals. Many of the texts (from all sources) are based on the authors' personal experiences, revealing that Talian was their childhood language.[5]

While articles from *O Florense* are available online in a format that can be copied and pasted into a text editor, articles from *Correio Riograndense* are available online as images, and thus need to be digitized to be analyzed. In the case of book excerpts, none of the books that have been included in the corpus are available online. As a result, these excerpts need to be scanned and subsequently digitized.

Once the book excerpts and articles from *Correio Riograndense* are collected, we use OCR (optical character recognition) to turn all images into editable text. In the case of *Correio Riograndense*, however, each

article page typically contains more than one article of interest as well as figures in them, and many key texts are originally printed in multiple columns, which is not appropriate for OCR given the presence of multiple breaks per line – see an example in Figure 2, where there are two texts in Talian, both of which have been included in the corpus (highlighted). Therefore, each article page needs to be prepared prior to OCR. The preparation involves (i) removing any non-text objects from the page, (ii) extracting all articles on the page to save them as individual files, and (iii) verticalizing texts spanning over multiple columns (e.g. *Difesa del regno* in Figure 2). Once each article is its own separate image file, we OCR all files using Tesseract (Smith 2007) in R (R Core Team 2023).

CORREIO RIOGRANDENSE - Caxias do Sul, 17 de março de 2010

"Che dona, che vipera, come zelo che sto poro Dio el pol suportar?
Qua ghe manca na man. Olo de rissino e manganelo!"
Nanetto Pipeta

17 - Imigração

Vita, Stòria e Fròtole

El ritorno de Nanetto Pipetta (555)

I giughi de carte

Geraldo Sostizzo
Agente Consular Italiano, Cascavel - PR

Chi no ga giugà carte quando el gera picolo? El giugo dele carte fea parte del divertimento dei òmeni. Ghe gera anca done che dele volte ghe piasea giugar carte.

Gaveimo diversi tipi de giughi: el pi comun gera la *briscola* che se giughea in due o in quatro, gera el giugo pi ficile. Gaveimo anca el giugo de *scova*, anca questo se giughea in quatro o anca in due e bisognea che le carte sora la tola le somasse 15, lora se podea catarle su e se tute somea 15, lora se fea na *scova*. che valea un punto.

Nantro giugo gera el *trisete*, anca questo se podea giugarlo in due o in quatro. Altro gera el quatrilio. Questo bisognea giugarlo in quatro e anca se podea giugarlo in cinque che se ciamea *cinquilio*, sol che ogni giugada, un stea fora, i vol dir che sempre se giughea in quatro.

Nantro giugo zera el *meno cope*. Vol dir che guadagnea chi fea meno carte de cope.

El barato gera fomà de quatro nàpole: la nàpola de oro, de cope, de bastuni e de spade. Ogni nàpola gavea diese carte: el asso, el due, el trè, el quatro, el cinque, el sei, el sete, el fantin, el caval e el rè. La carta pi importante gera sempre el asso, dopo vegnea el trè, dopo el due, dopo el rè, el caval, el fantin, el sete, el sei, el cinque e la carta meno importante zera el quatro. Quando tel giugo de *trisete* un acusea na nàpola, volea dir che el gavea in man: el asso, el due e el trè dela stessa nàpola.

Tuti questi giughi, se giughea col barato spagnol. Ghe gera anca altri tipi de giughi, che pi tardi i ga scominsià giugar come el *pife*, el *pontin*, la *canastra*, el *nove* e altri giughi. Tuti questi giughi con nantro barato che gavea novanta carte, diverso de quel spagnol che ghinavea quaranta.

Le carte se giughea sol al sabo dopo mesogiorno, de note o a la diménega dopo dela messa o del rosàrio.

I veci giughea la diménega intiera e chi perdea, se tanto, perdea due o trè fiorini, lora stea atento par torli indrio tel prósimo giugno. I picoli no podea mia giugar tela bodega, sol casa coi fradei e el papà, sensa far bacan anca. Ma vegnea fora tante barufe par motivo dele carte. Ghe gera sempre qualcuno pi furbo che volea ciavar i altri, lora se lo bastonea e finia el giugo.

Difesa del regno

Mario Gardelin
Caxias do Sul (RS)

(Retomado do texto de Mario Gardelin, interrompido na edição de 7 de outubro de 2009).

Co'l se ga nicorto, Júpiter se ga spaventà. I gigantoni i zera rivai dal Cabo da Boa Esperança, ai pié del Olimpo. Come i gavesse fato, no'l savea. El zera sicuro che no i zera amighi.

Bisognava pareciar la difesa. Júpiter ga fato sonar le trombe, ciamando i dei a difender la casa comun. Come zera tanti ani che no i gueriava, i pensava che ghe saria un balo. I ze vegnesti fora co bele falde, profumai; le done, lora le zera bellissime. I giganti i ga spalancà i oci. Saria un visior? Lora, i ga sgnairà i so spadoni. E i se ga fato avanti, sburegando... Le dee e i dei dal spavento i se ga lassà cader in tera. Ma, tra de lori ghin zera anca de quei che i gavea conservà el vècio coraio. Pochi, ma disposti.

I gigantoni, co li ga visti, i se ga messi a rider. E i disea:

- Andè a dormir che ze meio! Sio strachi de esser dei? Quanti ani gaveo? Sei mila?

I valenti i rispondea:

- Semo de poche cilcole. E ve femo sentir le nostre spade. E i gigantoni i rispondea:

- E noantri ve femo sentir le nostre sassade.

Dei e gigantoni i ze partii al cumbate. I gigantoni i ga impienà le man de sassi e via! In pochi minuti i dei i zera tuti par tera, svanii e insenriniti dai colpi in testa. Ma, no i se lamentava. Júpiter, che'l se gavea rampegà fin al alto dela colona, el ga messo le man sui oci e el ga ditto che no ghe zera altro da far. Almanco se gavesse qua la me Juno, son sicuro che la me giutaria.

El aiuto ze vegnesto. Tuti savemo chi sia stato el gigante Golia, quel che Davide, futuro re d'Israele, ga copà co na fiondada. Ben, el gavea ledesto:

- Salve, Júpiter, belo e caro, come na spòvisina, che cade su na galina; Salve deo dei, dei del Olimpo, dele man e dei cei; Teror dei gigantoni, dei gatini e dei gatoni!

- Fermo! Fermo! Ga ditto Juno. I gigantoni chi li ga spaurà son stà mi. Se taso, par piasserghe al me omo, par el culto dela personalità, i ghe dà veri e falsi onori. E el me omo no'l schersa mia. El aceta tuto, e el ghin inventa nantro milion par sora.

Nanetto ze restà mararveià:

- Che dona, che vipera, come zelo che sto poro Dio el pol suportar? Qua ghe manca na man. Olo de rissino e manganelo!

- Nanetto, tento! Mi ledo i to pensieri. Métete a posto, se no te vol sentire le curesse de un sicoto su la schena.

- Cara signora, vui dir: cara Dea, bela dona.

- Nanetto, dise Juno, go sempre pensà che te sì un parsimònio! Vedo che te sì inteligente!

E Nanetto, lora, el ga pensà de remediar la situassión e el ga continuà la fròtola:

- Júpiter, deo glorióso; quando el resta furioso, treme le capuere e anca el s-cioso. Júpiter, sempre zóvane, che porta ben, un milion de ani, patatin, patatan... E lì Nanetto se ga imbalbì:

- Ciao, Júpiter; stame ben; tento co la to dona; la te spaca i sarvei!

E de un sulto, el se ga rampegà in gropa al cavalin Pirolipirolipirolin. E el ga osà in brasilier, par farghe un dispeto a Jupiter:

- Velho bobo, deus batata, não vale uma lata. não vale um tostão, nem uma garfada de feijão. Júpiter no'l ga capio gnente e el ga domandà:

- Cosa galo dito sto tàngaro? - E Juno, svelta, la ga taià:

- El ga fato un ditirumbo in onor dele done de tuti i tempi, de tuti i posti, vere benedission de Dio. Sensa la dona, el omo saria un maledìo, un caicio storto!

Etnias

Projeto de Frei Rovílio Costa *[in memoriam]*, em parceria com CR e revista Insieme

O italiano que esta em mim

Fábio Chiamenti
Professor, Erechim (RS)

Quarto filho de Benvenuto Chiamenti, nasci em 29 de julho de 1929 em Nonoai - onde também nasceram meus seis irmãos. Aos seis anos, meus pais se transferiram para Boa Vista, atual Erechim. Feitos os primeiros estudos em Erechim, conclui-os no Seminário de Nossa Senhora da Saúde, em Marcelino Ramos. Fui estudar Filosofia na Pontifícia Universidade Gregoriana de Roma, onde passei sete anos. Voltando, soube que o Curso de Filosofia nessa Universidade não era reconhecido no Brasil - o refiz na Universidade de Passo Fundo. Minha profissão foi o magistério.

Pelo ano de 1975, fiz especialização em Estruturas do Português na PUCRS. Formei-me nesta mesma Universidade em Lingüística Aplicada. E passei a lecionar na Universidade de Erechim (URI). Antes de aposentar-me, fecionei na Escola de Idiomas Wizard.

Lembro do tempo da II Guerra Mundial. Falar italiano era proibido. Meu pai, Benvenuto, era italiano, filho de Michelangelo Chiamenti e de Lucia Deitos. Não conheci o meu avô; a vó Lucia era cega, devido à catarata, e morava conosco. Um pacto fora feito com o filho Benvenuto para cuidar dela enquanto vivia. Deixar-lhe-ia a quantia de dois contos de réis no banco, só retirável por sua morte. Assim, ela vivia conosco.

A ausência do vô Michelange-

lo, entendi em tempos mais recentes: morava em Caxambu do Sul e é certo que faleceu lá. Era o tempo da gripe espanhola no Brasil (1918...). Deve tê-lo ceifado como a tantos outros.

Minha mãe, Francisca de Assis Pereira de Campos, era brasileira. Ela, pela convivência com o marido, entendia bem o italiano. Nós, filhos, nunca falávamos italiano ou vêneto. Acontece, porém, que nessa época e circunstâncias, o italiano, assim como o alemão, eram reprimidos, e as pessoas dessas origens eram tachadas de "5ª Colunа". O vêneto a gente entendia um pouco, pois éramos rodeados de descendentes italianos. Mas entendia pouco mesmo. Tanto é verdade, que um dia meu pai me mandou: "Porta la égua par de sora del vignal". Isto é: leve a égua *pelo lado de cima* do parreiral. É eu a levei *debaixo* (desoto) do parreiral... Coitada! nem podia levantar a cabeça!

Mas, estudando em Roma, nossa casa internacional, cuja língua comum era o francês, aprendi o italiano oficial - língua bonita, clara e fácil de entender.

Nossa casa em Roma acolhia gente de diversas etnias. De sorte que num só dia falávamos francês em casa, italiano na rua, e latim na Pontifícia Universidade Gregoriana. Hoje, sou professor de italiano, francês, português e latim. (*Este artigo foi enviado há mais de dois anos. Fábio Chiamenti faleceu em 1º de março de 2010*).

Figure 2. A typical raw text sample from *Corrreio Riograndense* used in the Talian corpus. Articles in Talian are highlighted in red.

Tesseract is a cross-platform OCR engine created in the late 1980s by Hewlett-Packard. As of 2006, the engine is sponsored and developed by Google. Tesseract recognizes over 100 languages, and can be trained to include any language of interest. The engine can be used via command-line as well as through packages and libraries developed for different languages – e.g. R (tesseract) and Python (pytesseract). In our corpus, we use Tesseract both via command-line, in conjunction with ImageMagick for image preparation (The ImageMagick Development Team 2021), and via R with the tesseract package (Ooms 2021). In our case, we use trained data from Standard Italian, which offers the best orthographic approximation for Talian among the languages supported by Tesseract.

Once the texts are OCR'ed, they are manually checked for any processing errors. We then employ a series of R scripts to tokenize, syllabify, assign stress, and phonetically transcribe the data. These scripts take into account the phonotactic patterns of Talian as well as predictable phonological characteristics of the language. For example, stress in our corpus is assigned based on orthographic diacritics present in the original text, which indicate irregular stress, as well as the presence of a final heavy syllable (i.e. a final syllable ending in VV or VC), in which case stress is final.

Stress is not always predictable in our corpus, though, especially given that orthography is not standardized in Talian (and authors may not mark irregular stress in a consistent way). As a result, we estimate a conservative error rate of approximately 5% for stress assignment in the words in our corpus with our scripts. This estimate is based on ten separate samples of 100 random polysyllabic words each, which had an average error rate of 3.6% overall in stress assignment ($s = 1.8\%$).

Finally, once words are assigned stress, IPA transcription is generated with two separate scripts, which convert authors' observable orthographic conventions to phonetic symbols. As we will see below, the quality of mid vowels in the relevant data was checked manually. The corpus is then compiled using the RData format, which ensures better compression, metadata preservation, and better structural integrity. Figure 3 illustrates all the steps involved in the creation of our corpus, which at the time of this writing has 237,774 words. The corpus can be accessed from an Open Science Framework repository at <osf.io/63nrx>.

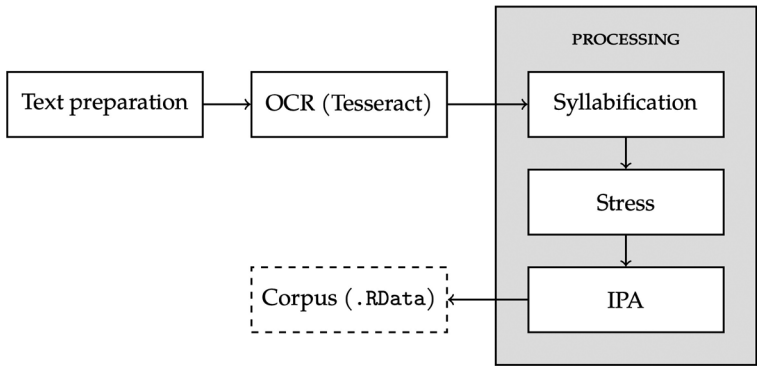

Figure 3. Steps involved in creating the Talian corpus. Final output is an RData file.

In the next section, we describe how the metaphony data were extracted from our corpus, as well as how these data were coded and analyzed.

5. Methods

The data analyzed in this article were extracted from the Talian Corpus, based on texts from 40 authors.[6] In order to obtain all the words that could undergo metaphony, we created an R script that identified all the words ending in /i/ with an underlying mid vowel in stressed position (which could be orthographically represented as {i u} or {e o}, that is, with or without metaphony). The script focused on polysyllabic words, since monosyllables that offer context for metaphony seem to be very limited in the language – or at least in our corpus, where only one monosyllabic lexical item seemed to be a reasonable candidate for metaphony (namely, [vui] 'want.1SG'). However, it is unclear whether this monosyllable in effect exhibits metaphony, as there are no alternations between high and mid vowel for it in the data.[7] As expected, there were no monosyllables with [ii] (see discussion in Section 3).

The polysyllabic words that were extracted from our corpus were manually checked by the second author, a heritage speaker of Talian, for quality of the stressed mid-vowel, given that it is not possible to predict whether orthographic {e o} represent low- or high-mid vowels. The use of diacritics is not of much help in this case – even though authors use diacritics relatively consistently to mark for irregular stress, these diacritics are not used with the same consistency to mark for vowel quality.

The verification of mid-vowel quality was also supported by the phonetic transcription in Luzzatto's (2000) Talian-Portuguese dictionary.

After mid-vowel verification, tokens with underlying /ɛ, ɔ/ were excluded. The total number of tokens with context for metaphony was 3085. Examples of target items are shown in (4) in orthographic form. Metaphony was implied for the items spelled with a stressed high vowel (in place of a high-mid vowel; see (4b)).

(4) Examples of target items (orthographic form)
a. Non-application of metaphony:
senti 'feel.2SG'
amori 'love.PL'
dóveni 'young.PL'
b. Application of metaphony:
curri 'run.2SG' (alternative form: *corri*)
cagniti 'dog.DIM.PL' (alternative form: *cagneti*)
fasui 'bean.PL' (alternative form: *fasoi*)

All the items were coded for application of metaphony (response variable) as well as for the following predictor variables: target vowel quality (/e/ or /o/), number of syllables in the word, morphology (whether the target vowel is in the root or in a suffix), onset and coda of the target syllable, and onset of the trigger syllable. In our analysis below, we focus specifically on words with penultimate stress with two or three syllables (n = 2091; 487 unique), given their representativeness in the data. As we will see in what follows, the predictor variables of statistical relevance here are number of syllables and morphology, and there is an asymmetric behavior between the target vowels, similarly to what has been observed with metaphony in final and pretonic position (see Section 3).

6. Results and analysis

Before we proceed to the analysis of metaphony in words with penultimate stress, we briefly examine the patterns that were obtained for target words with antepenultimate and final stress. There were 272 items in the corpus with antepenultimate stress which offered context for metaphony. In these items, the target vowel could be in antepenultimate and/or penultimate position (there were 191 items with target /e/ and 81 with target /o/). Metaphony with antepenultimate stress applied in 38 (14%) of these items. In the data, most cases of application with antepenultimate stress target the penultimate syllable (the word *òmini* 'man.PL' alone comprises 30 tokens). In the case of *òmini*, metaphony

is blocked in the stressed syllable as it exhibits a low-mid vowel (see (2)). Regarding other words with antepenultimate stress in the data, on the other hand, both the antepenultimate and the penultimate syllable could display metaphony, as they both have high-mid vowels underlyingly. However, metaphony targeting both the antepenultimate and the penultimate syllable is not observed in the data. What we find instead is metaphony targeting either the penultimate syllable (but not the antepenultimate; e.g. *giovini* 'young.PL') or the antepenultimate syllable (and skipping the penultimate; *dùveni* 'young.PL', *grústoli* 'type of pastry. PL'). The latter case appears to contrast with what has been observed in Central Veneto, where a high-mid vowel in posttonic position invariably raises when antepenultimate /e, o/ are targeted (Walker 2005).

There were 158 items with final stress in the corpus which offered context for metaphony (45 with /e/, 113 with /o/). As expected, metaphony is not observed with underlying stressed final /e/, as it would result in a [ii] string (see Section 3). With final stressed /o/, metaphony applies in 18 tokens (11.4% of the items with final stress). Examples of metaphony with final stress are *fasui* 'bean.PL', *pignui* 'pine seed.PL', *nissui* 'bedsheet.PL'.

Let us now turn to the items under focus in our analysis, namely words with penultimate stress and two or three syllables. In these items (n = 2091), metaphony is observed 18.3% of the time (i.e. in 383 tokens). Figure 4 shows the proportion of metaphony in these items. As we can see, metaphony is more frequent with /o/ than with /e/, with a potential interaction between target vowel and number of syllables in the word: whereas with stressed penultimate /e/ metaphony is more frequent in words with three syllables, with stressed penultimate /o/ metaphony seems to be slightly more frequent in words with two syllables.

To model the data, we ran a mixed-effects logistic regression with a by-author random intercept – as previously mentioned, given that no standard orthography exists for Talian, authors can potentially show different patterns of metaphony. Target vowel (targetV) and number of syllables (nSyl) were the predictors included in the model (main effects and interaction). Additional variables, such as segmental quality of onsets and codas (including their sonority), had no systematic effect in the data and were therefore not considered in our final model. The model estimates and associated 95% confidence intervals are provided in Figure 5, where the intercept represents two-syllable words with target vowel /o/.

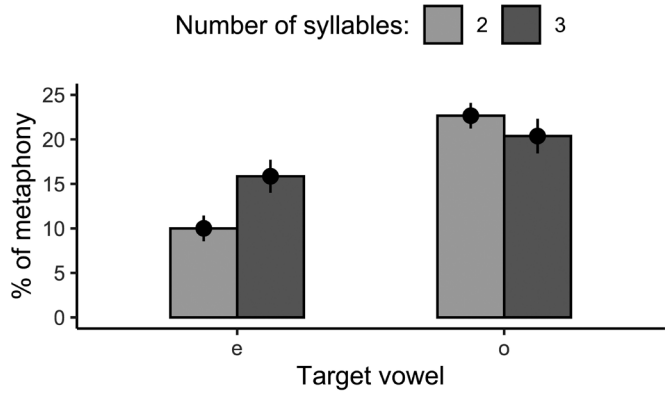

Figure 4. Proportion of metaphony by target vowel in two- and three-syllable words.

In Figure 5, effect sizes ($\hat{\beta}$) are provided in log-odds – by-author random intercepts are not shown ($s^2 = 2.1$). Negative estimates indicate a lower probability of metaphony. For example, for targetV [e], $\hat{\beta} = -1.26$,95% $CI = [-1.68, -0.86]$. This effect can be interpreted as follows: for two-syllable words (our reference level in the model), metaphony with /e/ is less likely than metaphony with /o/. nSyl [3] indicates that there is no statistically significant effect of number of syllables when the vowel is /o/, our reference level. If we look back at Figure 4, this result is unsurprising, given the standard errors observed for /o/. Finally, and more importantly, the interaction between number of syllables and target vowel shows a significant effect: $\hat{\beta} = 1.01$,95% $CI = [0.42, 1.61]$. This, again, is not surprising given the patterns in Figure 4, since the effect of number of syllables is not constant across target vowels. For example, if we hold /e/ constant and go from two- to three-syllable words, we see a substantial increase in the percentage of metaphony in Figure 4. This is mirrored in our positive effect size in the model.

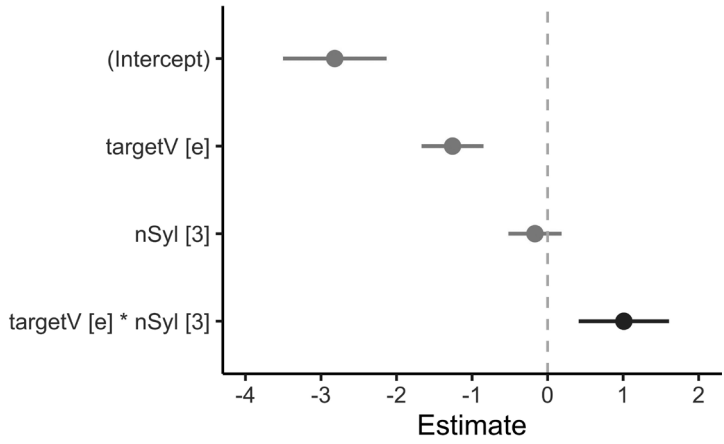

Figure 5. Estimates (log-odds) and associated 95% confidence intervals.
Model specification: *metaphony* ~ *targetV* ∗ *nSyl* + (1 | *author*)

Thus far, we can see an asymmetrical trend in our data: whereas for /e/ we see more metaphony in three-syllable words, the opposite seems to be the case for /o/, which displays more metaphony in two-syllable words, even though this trend is not statistically significant. This asymmetry was shown in our models' estimates in the significant interaction between number of syllables and target vowel. The question is what could be driving such an asymmetry. One potential factor involves lexical statistics. As shown in Figure 6, once we look at all two- and three-syllable words containing orthographic {e o} in penultimate position in our corpus (n = 54,748), we observe the same asymmetry: {e} appears more often in penultimate position in three-syllable words, while {o} appears more often in penultimate position in two-syllable words. These similarities between the rates of application of metaphony and the lexical distribution of target vowels is consistent with the idea that more frequent forms in one's lexicon are treated less faithfully (e.g. van Oostendorp 1997, Itô & Mester 2001, Coetzee & Kawahara 2013). If that is indeed the case, the authors of our texts are applying metaphony more often to patterns which are more frequent in their lexica.

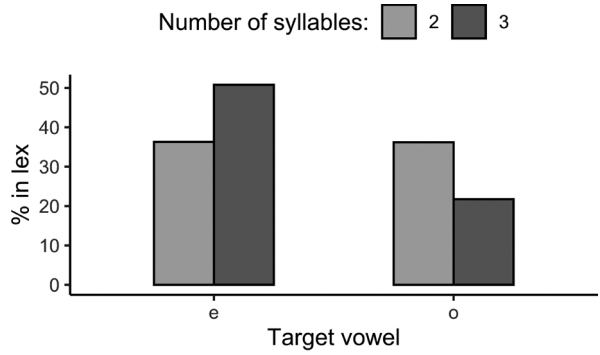

Figure 6. Frequency of {e o} in penultimate position in the corpus (n = 54,748).

Next, we look at morphology: are target vowels more susceptible to metaphony when they are in the root, as opposed to when they are in a suffix? As previously mentioned, we coded all the target vowels as 0 (not in root) or 1 (in root). Figure 7 shows that, for morphology, we find an asymmetrical pattern once again: whereas for /e/ we observe more metaphony when the target vowel is not in the root (e.g. *cagn-**i**t-i* 'dog.DIM.PL', the target vowel is bolded), the opposite is true for /o/ (e.g. *col**u**r-i* 'color.PL'). These trends are statistically confirmed by a mixed-effects logistic regression – estimates are shown in Figure 8.

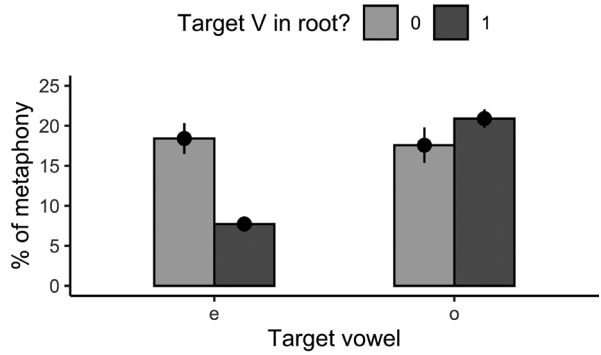

Figure 7. Proportion of metaphony by target vowel by presence of target vowel in root.

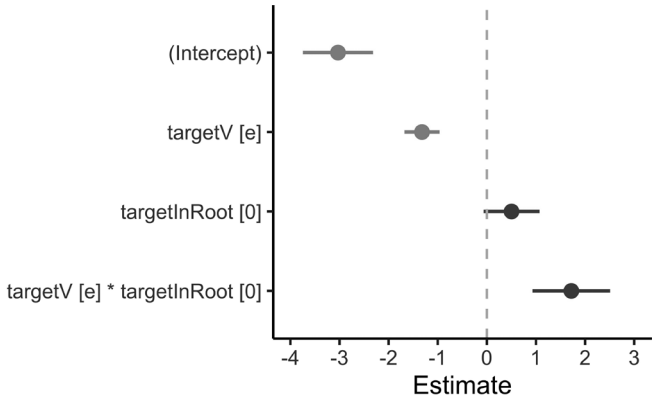

Figure 8. Estimates and associated 95% confidence intervals.
Model specification: *metaphony* ~ *targetV* ∗ *targetInRoot* + (1 | *author*)

In summary, our corpus data point to a target vowel asymmetry in Talian metaphony observed in penultimate syllables, which is constrained by number of syllables in the word and morphology (see Table 2). If this asymmetry is present in speakers' grammars, the question is how we can model such patterns in a way that takes into account the potential effect of lexical statistics.

	/e/	/o/
Entire corpus	more frequent in 3-syl words	more frequent in 2-syl words
Target items	more metaphony in 3-syl words	more metaphony in 2-syl words
Vowel in root	less metaphony	more metaphony

Table 2. Summary of results for penultimate /e, o/.

7. Formalizing metaphony in Veneto

To formalize our results, we need a framework that allows for the variable application of metaphony. Constraint-based probabilistic frameworks such as Maximum Entropy Grammar (e.g. Goldwater & Johnson 2003; Wilson 2006; Hayes & Wilson 2008) can easily accomplish such a task. In a MaxEnt grammar, constraints are weighted in such a way that the cost of violating constraint C is proportional to the weight of said constraint, i.e. w_C.

In addition to a probabilistic framework such as MaxEnt, we also need some form of 'lexical regulation' to capture the potential effect of lexical statistics discussed above. We follow Coetzee & Kawahara (2013) and assume that faithfulness constraints have a scaling factor that perturbs their weight on the basis of the lexical distribution of a relevant pattern. For example, if pattern *P* is more frequent in the lexicon, then the weight of faithfulness constraints for *P* are scaled down in such a way that changing *P* is less costly in the grammar. This captures the generalization that higher frequency tends to correlate with less faithfulness on the surface.

Finally, we need to motivate metaphony itself in the grammar. We follow Walker (2005, 2010), who proposes that a [+high] feature in a posttonic syllable must be associated to the stressed syllable. This can be captured with a licensing constraint, as discussed below. Figure 9 illustrates feature licensing in Talian metaphony.

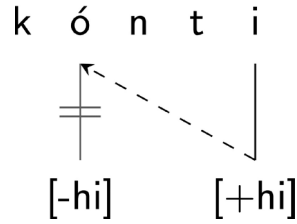

Figure 9. [+hi] licensing resulting in metaphony in Talian.

To illustrate how we can formalize the asymmetry in the application of metaphony with target /e, o/ in penultimate position in two- and three-syllable words, our grammar will rely on the following constraints. IDENT[hi] penalizes divergences in vowel height (that is, involving the feature [+high]) between the input and the output. As we will see below, this constraint can be scaled in order to reflect the differences observed between two- and three-syllable words. LICENSE([+hi]$_{posttonic}$, *ó*), on the other hand, ensures that a stressed high-mid vowel will be raised to [+high]. Contrary to Walker (2005, 2010), we assume that licensing involves the unstressed final vowel only, rather than any posttonic vowels, since we have no evidence in our data that metaphony is triggered by /i/ in other posttonic positions. Finally, OCP-V (Obligatory Contour Principle; Leben 1973) penalizes identical vowels in adjacent syllables. The consequence of this constraint is that metaphony with /e/ will be less favored than metaphony with /o/, given that the phenomenon is invariably triggered by word-final /i/. This is

one way to account for the overall higher rates of metaphony with target /o/ in Figure 4 – it is also well motivated, given that less than 14% of words in our corpus contain identical vowels in adjacent syllables within the stress domain. These constraints are summarized in (5).

(5) **Constraints**[8]

IDENT[hi]:	every [high] in the output must have a correspondent in the input
LICENSE([+hi]$_{posttonic}$, σ):	the unstressed final [+high] must be associated to the stressed syllable
OCP-V:	vowels in adjacent syllables must not be identical

Our present analysis focuses solely on the target vowel asymmetry between penultimate /e, o/ observed in words with two and three syllables. It does not include the asymmetry between /e, o/ regarding morphology. As discussed in the previous section, /e/ undergoes metaphony more frequently when it is outside the root, while /o/ undergoes the process more frequently when it is in the root. We contend that this pattern can be accounted for by a weighted constraint such as IDENT$_{morph}$, which penalizes differences in input-output correspondence for vowels inside and outside the root.

Turning to the patterns under analysis, Tableaux 1-4 illustrate how two- and three- syllable words with a target vowel in penultimate position are evaluated. At the top of each tableau, we see the weight of each constraint. Superscript '+' indicates that the weight for IDENT[hi] will be scaled on the basis of lexical statistics for a given word/pattern. The column H(x) simply sums over all the violations of a given candidate (multiplied by the weight of said violations). P*(x) is our MaxEnt Score, which is calculated by exponentiating the negated value given in H(x). Finally, P(x) simply calculates the probability for a candidate to surface – note that this number depends on the size of the set of candidates being evaluated. For convenience, column R indicates the real proportion of application/non-application of metaphony found in our data. Since the application of metaphony is dependent on stress, the position of stress is indicated in the input below.[9]

The scaling factors assumed here are: e-i$_{2syl}$ = 1, e-i$_{3syl}$ = 0.4, o-u$_{2syl}$ = 1, and o-u$_{3syl}$ = 1.7. These are illustrative factors in order to capture the empirical patterns we observe. Ultimately, scaling factors should be interpreted relative to the constraint weights they modulate, just like constraint weights should be interpreted relative to each other, not in isolation. For example, for a two-syllable input such as /ˈpes-i/ 'fish.PL', in Tableau 1, we must add 1 to the weight of IDENT[hi]. In Tableau 2, on the other hand, where we have a three-syllable input (/kalˈset-i/ 'sock.PL'), we add 0.4 to the weight of IDENT[hi]. In both tableaux, we can easily generate probabilities (P(x)) that approximate what we

observe in the data (R). Tableaux 3 and 4 demonstrate the same analysis for target /o/. Notice that the scaling factors for three-syllable words are different between e-i and o-u, which is necessary to capture the asymmetry found in the data.

'fish.PL'	$w = 1^{+1}$	$w = 1$	$w = 0.8$				
/ˈpes-i/	IDENT	OCP	LIC[+hi]	H(x)	P*(x)	P(x)	R
[ˈpesi]			1	0.8	0.449	0.90	0.898
[ˈpisi]	1	1		3	0.05	0.10	0.102

Tableau 1. Metaphony with target /e/ in penultimate position in two-syllable words.

'sock.PL'	$w = 1^{+0.4}$	$w = 1$	$w = 0.8$				
/kalˈset-i/	IDENT	OCP	LIC[+hi]	H(x)	P*(x)	P(x)	R
[kalˈseti]			1	0.8	0.449	0.83	0.837
[kalˈsiti]	1	1		2.4	0.091	0.17	0.163

Tableau 2. Metaphony with target /e/ in penultimate position in three-syllable words.

'buck.PL'	$w = 1^{+1}$	$w = 1$	$w = 0.8$				
/ˈkont-i/	IDENT	OCP	LIC[+hi]	H(x)	P*(x)	P(x)	R
[ˈkonti]			1	0.8	0.449	0.77	0.773
[ˈkunti]	1			2	0.135	0.23	0.227

Tableau 3. Metaphony with target /o/ in penultimate position in two-syllable words.

'color.PL'	$w = 1^{+1.7}$	$w = 1$	$w = 0.8$				
/koˈlor-i/	IDENT	OCP	LIC[+hi]	H(x)	P*(x)	P(x)	R
[koˈlori]			1	0.8	0.449	0.87	0.796
[koˈluri]	1			2.7	0.067	0.13	0.204

Tableau 4. Metaphony with target /o/ in penultimate position in three-syllable words.

As Tableaux 1-4 indicate, the patterns for metaphony in two- and three-syllable words with penultimate stress in Talian can be captured with IDENT[hi], OCP and LIC([+hi]$_{posttonic}$, *σ́*) as weighted, scalable constraints.

8. Final remarks

By examining a large corpus of written Talian, we uncovered a target vowel asymmetry in the application of metaphony. Because Talian is an understudied language and lacks standardized orthography, our assumption was that written data might be a proxy for the grammar of the language, insofar as authors whose texts contain higher rates of orthographic metaphony likely show higher rates of metaphony in their speech. The asymmetry in question is manifested in two ways: (i) target /e, o/ exhibit distinct patterns of application in penultimate position in two- and three-syllable words, and (ii) target /e, o/ exhibit distinct patterns of application depending on whether the vowel is in the root or in a suffix.

We hypothesized that the asymmetry involving number of syllables in the word may be driven by lexical statistics, given that the distribution of target vowels in the entire corpus matches the patterns obtained for metaphony. In other words, metaphony with a given target vowel applies more frequently in positions where that target vowel is more frequent in the language. This observation is based on the cross-linguistic finding that forms which are more frequent tend to be produced less faithfully (van Oostendorp 1997, Itô & Mester 2001, Coetzee & Kawahara 2013).

Finally, we illustrated how a constraint-based grammar can account for the asymmetry in question as well as lexically-specific effects by employing a scaling factor, which perturbs the weight of faithfulness constraints. The scaling factor in our analysis is associated to an identity constraint (IDENT[hi]), which penalizes input-output height mismatches, on the basis of the number of syllables in the word. Other constraints included in the analysis are OCP-V, which accounts for metaphony applying overall more frequently with target /o/ than target /e/, and LIC([+hi]$_{posttonic}$, *σ́*), which requires that the [+high] feature of an unstressed final vowel be associated to the stressed vowel.

Although we assume that the orthographic patterns may reflect authors' productions, we cannot confirm whether the asymmetries we unveiled are part of speakers' grammars without an analysis focusing on metaphony in Talian speech. A central question left for future research is thus whether speakers of Talian in effect mirror the patterns in our

corpus. Future research should also be able to determine which linguistic factors (segmental, prosodic and/or morphological) condition metaphony in speakers' speech, whether the process is conditioned by any social factors, and how pervasive it is in Talian-speaking communities. These observations should contribute to establishing further comparisons between Talian and Central Veneto.

Abbreviations

IIA = Italian Immigration Area; MaxEnt = Maximum Entropy; nSyl = number of syllables; OCP = Obligatory Contour Principle; OCR = Optical Character Recognition; P = pattern; targetV = target vowel; w_C = weight of constraint.

Acknowledgements

We would like to thank the following research assistants for their help preparing the corpus files: Émilie Dubé, Alexandra Lancaster, Ray Marks, Fabian McCarthy, Lamia Oudni, Carolyn Rathgeber, and Jovia Wong (McGill University); Jenna Gramlich (Ball State University); and Hannah Markert (Saint Mary's University). We are thankful to Valdemir Guzzo for his assistance in finding corpus materials as well as more information on the history of the Italian Immigration Area. We also thank two anonymous reviewers for their feedback, as well as editors Margherita Di Salvo and Eugenio Goria.

Notes

[1] Although Talian has been recognized as an official language by many municipalities in the IIA and elsewhere, it has in effect the status of a heritage language in these communities – the language is predominantly spoken at home or in tightly-knit social circles, and it is not used for formal instruction, official business deals nor governmental affairs (see e.g. Pertile 2009). In general, speakers who report using Talian on a regular basis live in rural areas and/or are middle-aged or older (e.g. Pertile 2009; Guzzo & Garcia 2020).

[2] Talian has also been influenced by contact with Brazilian Portuguese. With respect to phonology, for example, some borrowings from Portuguese may variably exhibit /ʃ, ʒ/, which are absent from the Veneto inventories (Margotti 2004). At the same time, Talian has also influenced the Portuguese varieties with which it is in contact. For example, speakers of these Portuguese varieties exhibit lower rates of vowel reduction in unstressed syllables (see e.g. Guzzo & Garcia 2020), as well as rhotic substitution (/ɾ/ instead of /r/, since Veneto varieties have a single rhotic phoneme; see e.g. Battisti & Bovo 2004; Guzzo *to appear*).

[3] Following Guzzo (2022), we assume that words like /fa'zoi/ and /ni'soi/ exhibit VV strings. These strings may be realized on the surface as diphthongs, which supports their classification in (1) as having final stress. As discussed in what follows, items with this profile were discarded from the analysis, given that (a) they are relatively infrequent in the data, and (b) metaphony targets word-final /oi/ but not word-final /ei/.

[4] *Correio Riograndense* changed its name to CR4 in 2017, when the newspaper discontinued its print editions. Its online editions can be accessed at <www.tuaradio.com.br>. The articles from *Correio Riograndense* that are included in the corpus were obtained through a search on the database of Câmara Municipal de Vereadores de Caxias do Sul ('Caxias do Sul City Council'), which contains the archives of regional newspapers (available at <liquid.camaracaxias.rs.gov.br/portalliquid>). The articles from *O Florense* are all part of a monthly column called *Ciàcole* 'chitchat', which can be accessed on <www.jornaloflorense.com.br/colunistas/toni-sbrontolon/16>.

[5] Editorial adjustments seem to be kept to a minimum in *Correio Riograndense,* which is our only source to publish many different authors. This is evidenced by the fact that a single word may be spelled in different ways (across authors or by the same author), often mirroring the different ways in which it can be pronounced.

[6] One of the texts in *Correio Riograndense* is not specified for author. This text was labeled as having an unknown author in the corpus and was included in the analysis.

[7] In a Talian grammar (Stawinski, 1982), *vúi* is provided as an alternative form for *vóio*, but not for *vói*.

[8] The highly-ranked constraint IDENT_{IO}(ATR), which penalizes metaphony targeting vowels other than /e, o/, is implied in this analysis (Walker 2010).

[9] Stress assignment is beyond the scope of the present paper. Therefore, we do not include constraints responsible for assigning stress in our analysis.

References

Battisti, Elisa & Bovo, Nínive Magdiel Peter 2004. Variação linguística como prática social: Análise quantitativa e qualitativa da realização da vibrante no português em contato com italiano [Language variation as social practice: Quantitative and qualitative analysis of vibrant sounds in Portuguese in contact with Italian dialects]. *Lingua(gem)* 1,2. 107-123.

Belloni, Silvano 2009. *Grammatica veneta*. Padova: Esedra.

Coetzee, Andries W. & Kawahara, Shigeto 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31,1. 47-89. <www.jstor.org/stable/42629730>.

De Boni, Luís A. & Costa, Rovílio 1979. *Os italianos do Rio Grande do Sul* [The Italians from Rio Grande do Sul]. Porto Alegre: EST; Caxias do Sul: EDUCS.

Eisenstein, Jacob 2015. Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics* 19,2. 161-188. <doi.org/10.1111/josl.12119>.

Frasson, Alberto 2021. Clitics are not enough: on agreement and null subjects in Brazilian Venetan. *Glossa: A journal of general linguistics* 6,1. 86. <doi.org/10.5334/gjgl.1697>.

Frosi, Vitalina M. & Mioranza, Ciro 1983. *Dialetos italianos: Um perfil linguístico dos ítalo-brasileiros do Nordeste do Rio Grande do Sul.* [Italian dialects: A linguistic profile of Italo-Brazilians in northeastern Rio Grande do Sul]. Caxias do Sul: EDUCS.

Frosi, Vitalina M. & Mioranza, Ciro 2009. *Imigração italiana no nordeste do Rio Grande do Sul* [Italian immigration in northeastern Rio Grande do Sul]. 2nd edition. Caxias do Sul: EDUCS.

Goldwater, Sharon & Johnson, Mark 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*. 111-120.

Guzzo, Natália Brambatti 2022. Brazilian Veneto (Talian). *Journal of the International Phonetic Association* (Illustrations of the IPA). <doi.org/10.1017/S002510032200010X>.

Guzzo, Natália Brambatti *to appear*. Prosodically-conditioned variation: Rhotics in Brazilian Veneto. In Rao, Rajiv (ed.), *The Phonetics and Phonology of Heritage Languages*. Cambridge: Cambridge University Press.

Guzzo, Natália Brambatti & Garcia, Guilherme D. 2020. Phonological variation and prosodic representation: Clitics in Portuguese-Veneto contact. *Journal of Language Contact* 13,2. 389-427. <doi.org/10.1163/19552629-bja10021>.

Hayes, Bruce & Wilson, Colin 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39,3. 379-440. <doi.org/10.1162/ling.2008.39.3.379>.

Itô, Junko & Mester, Armin 2001. Covert generalizations in Optimality Theory: The role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology* 7. 273-299.

Loriato, Sarah 2019. Language use and intergenerational transmission of heritage Veneto in the rural area of Santa Teresa, Brazil. *International Journal of the Sociology of Language* 260. 37-59. <doi.org/10.1515/ijsl-2019-2047>.

Luzzatto, Darcy Loss 2000. *Dissionàrio talian (véneto brasilian)-portoghese* [Talian (Brazilian Veneto)-Portuguese dictionary]. Porto Alegre: Sagra Luzzatto.

Maiden, Martin 1987. New perspectives on the genesis of Italian metaphony. *Transactions of the Philological Society* 85. 38-73.

Margotti, Felício Wessling 2004. *Difusão sócio-geográfica do português em contato com o italiano no sul do Brasil* [Socio-geographic diffusion of Portuguese in contact with Italian in southern Brazil]. PhD dissertation, Universidade Federal do Rio Grande do Sul.

Ooms, Jeroen 2021. tesseract. R package version 4.1. <CRAN.R-project.org/package=tesseract>.

Perrone, Alessia 2016. *Il processo della metafonesi nei dialetti italiani*. Master's thesis, Università degli Studi di Padova.

Pertile, Marley Terezinha 2009. *O talian entre o italiano padrão e o português brasileiro: manutenção e substituição linguística no Alto Uruguai gaúcho* [Talian between Standard Italian and Brazilian Portuguese: linguistic preservation and substitution in the southern region of Alto Uruguai]. PhD dissertation, Universidade Federal do Rio Grande do Sul.

R Core Team 2023. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Savoia, Leonardo & Maiden, Martin 1997. Metaphony. In Maiden, Martin & Parry, Mair (eds.), *The dialects of Italy*. London: Routledge. 15-25.

Smith, Ray 2007. An overview of the tesseract OCR engine. *Ninth international conference on document analysis and recognition (ICDAR 2007)* 2. 629-633.

Stawinski, Alberto Victor 1982. Gramática e vocabulário do dialeto italiano riograndense [Grammar and vocabulary of the Italian dialect from Rio Grande do Sul]. 3rd edition. In Bernardi, Aquiles, *Vita e stória de Nanetto Pipetta: nassuo in Itália e vegnudo in Mérica par catare la cucagna* [Life and history of Nanetto Pipetta: born in Italy and moved to America to search a dream]. 7th edition. Porto Alegre: EST; Caxias do Sul: EDUCS.

The ImageMagick Development Team 2021. ImageMagick. Retrieved from <imagemagick.org>.

van Oostendorp, Marc 1997. Style levels in conflict resolution. In Hinskens, Frans; van Hout, Roeland & Wetzels, Leo (eds.), *Variation, change and phonological theory*. Amsterdam: John Benjamins. 207-229.

Walker, Rachel 2005. Weak triggers in vowel harmony. *Natural Language and Linguistic Theory* 23. 917-989. <doi.org/10.1007/s11049-004-4562-z>.

Walker, Rachel 2010. Nonmyopic harmony and the nature of derivations. *Linguistic Inquiry* 41. 169-179. <www.jstor.org/stable/40606834>.

Wickham, Hadley 2014. Tidy data. *Journal of Statistical Software* 59,10. 1-23. <doi.org/10.18637/jss.v059.i10>.

Wilson, Colin 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30,5. 945-982. <doi.org/10.1207/s15516709cog0000_89>.

Zamboni, Alberto 1974. *I dialetti del Veneto*. Pisa: Pacini.