Form, function, and reduction: Phonetic erosion in the grammaticalization of Brazilian Portuguese *tipo*

Abstract

Grammaticalization of the noun *tipo* 'type, kind, sort (of)', including its use as a discourse marker, has been documented for several decades in Brazilian Portuguese in Brazilian Portuguese. Several processes have been shown to take place as linguistic elements acquire new syntactic and discursive functions, including reduction or erosion of phonetic and phonological structure. Review of the literature suggests that erosion is often analyzed as an an overall durational reduction while processes that may be taking place at the segment level are rarely investigated. This study aims to contribute to the current literature by investigating acoustic production of *tipo* among a group of middle-school students in Rio de Janeiro. More specifically, we document several forms and functions of use of *tipo* and examine phonetic differences at the segment level. Our results demonstrate that there are significant and consistent/regular differences between nominal and non-nominal (i.e., grammaticalized) usages of *tipo*—while some of these differences can be described as reduction processes, others defy such characterization.

Keywords: grammaticalization, erosion, Brazilian Portuguese, innovation, vernacular

1 Introduction

The noun *tipo* 'type, kind, sort (of)' has been undergoing grammaticalization in Brazilian Portuguese for several decades (Author(s), 2019; Bittencourt, 1999; Laurentino, 2016; Lima-Hernandes, 2005a, 2005b). Its grammatical functions are varied and known to include use as a comparative marker and a discourse marker, among others. The acquisition of grammatical functions by *tipo* has parallels among similar, sometimes cognate, taxonomic nouns in a variety of Indoeuropean languages, including: French (Cheshire & Secova, 2018), Italian (Voghera, 2013), Russian (Kolyaseva & Davidse, 2018), Spanish (Fernández, 2017), Swedish (Rosenkvist & Skärlund, 2013), and English (Traugott, 2008).

As a grammaticalizing form, one area of interest concerns the phonetic production of *tipo*, as grammaticalization processes are often correlated with phonetic erosion or reduction (Bybee, 2003; Heine, 2003, 2014; Meillet, 1912). Our study investigates *tipo* with consideration to its wide range of forms and uses, and considers erosion or reduction as a complex process which may accompany grammaticalization, but without assuming that all functions and forms should be impacted equally or in the same manner. To capture what appears to be a dynamic, ongoing change, we investigate production of *tipo* within a corpus of recorded interviews conducted among middle-school children in a lower socioeconomic region of Rio de Janeiro, Brazil. As we will show, our results indicate that while grammaticalization does indeed impact the phonetic structure of *tipo*, this varies according to its particular function, and is not associated with wholesale erosion/reduction across all segments and syllables. Instead, the phonetic developments taking place as part of the grammaticalization of *tipo* are shown to be more nuanced when aspects of individual segmental and sub-segmental productions are taken into account, and are not entirely characterizable as reduction or erosion.

1.1 The grammaticalization of *tipo*

Tipo, a masculine noun in Portuguese meaning 'kind, type, class, model' (cognate with English *type*) has, in recent decades, been used in a variety of grammatical functions which depart from its canonical use as a noun. Drawing on data collected in Rio de Janeiro during the 1970s as part of the NURC-RJ corpus of spoken interviews (Callou, 1991), de Castilho (1991) discussed functional changes of *tipo* occurring in educated Carioca (demonym for residents of Rio de Janeiro) speech. The first researcher to fully recognize these and other usages of *tipo* as grammaticalizations was Bittencourt (1999), who posited that possible innovation in its use, at times alongside the adverb *assim* 'this way/manner', had resulted in an expression that means 'something of this/that type'. Bittencourt further described how, through a process of semantic bleaching and change, *tipo* might be acquiring multiple functions beyond that of its original nominal meaning, a prediction borne out by subsequent research. Lima-Hernandes (2005a, 2005b) investigated data from the NURC-RJ corpus as well as the later PEUL spoken interview corpus (Scherre & Naro, 1992) and a combination of contemporary linguistic speech corpora and printed documents from the 13th through 20th centuries. Through synchronic and diachronic analyses she identified *tipo* as performing five grammatical functions, including its original nominal function. All subsequent work on Carioca usage of *tipo* since Lima-Hernandes (e.g., Thompson, 2013, 2019), has derived from the corpora mentioned above, meaning that the present study is the first substantive update to our knowledge of *tipo* usage in contemporary Carioca spoken discourse since the beginning of the 21st century.

While the focus of study for *tipo* grammaticalization has largely concerned the Carioca dialect, it has also been investigated in some other dialects of Brazilian Portuguese. Bertozzo (2014) examined the functions of *tipo* in the city of Chapecó, in the south of Brazil, while Laurentino (2016) investigated the functions of *tipo* in the Natalense dialect of Natal, in Brazil's northeast and geographically distant from both Rio de Janeiro and Chapecó. There do not appear to be large or substantive differences between these dialects

in terms of the grammatical functions for which *tipo* is employed. While researchers differ considerably in terms of description and naming of its various functions, there is general agreement on its use as a comparative marker and as a discourse marker or particle (see Author(s), 2021, for a more complete overview of the literature on the grammaticalized functions of *tipo*). Finally, to our knowledge no existing studies of *tipo* have investigated its phonetic properties.

1.2 The social context of Ilha do Governador

In 2015, we began fieldwork on what would eventually become the *Projeto Sociolingüístico Contemporâneo Brasileiro* 'Contemporary Brazilian Sociolinguistic Project' (Author(s), 2016, 2019) or PSCB corpus. The aim of the PSCB is to document innovations in Brazilian Portuguese which are particularly prevalent in, although not unique to, the speech of youth in Rio de Janeiro. Among these innovations is the use of *tipo* to serve a variety of grammatical functions, which forms the focus of this study. Recruitment and data collections sessions for the PSCB took place at two public schools (see section 2.1) on *Ilha do Governador* 'Governor's Island', the largest island in Guanabara Bay which lies adjacent to Rio de Janeiro. Among its total population (2010) of over 212,000 inhabitants, thirty-two percent (over 67,000) of the island's residents live among its slums, known locally as *favelas* 'communities' (Cavallieri & Vial, 2012).

We made the choice to recruit at public schools on Ilha do Governador for several reasons. Ilha do Governador itself is somehwat geographically isolated from the rest of the city, with a mixed socioeconomic population. Public schools in the city are mostly attended by low and low-middle class students while most upper-middle- and upper-class students attend private schools. Public school students would, therefore, have relatively fewer opportunities to travel or otherwise be exposed to non-local speech patterns. In addition, the first author is a member of the community on the island, having attended municipal schools there; this both afforded greater access to the schools involved in the

study, and allowed her to serve as the interviewer for all of the recording sessions. A modified version of Labov (1984)'s sociolinguistic interview script, which encourages speakers to use informal language and speak as naturally as possible (D'Arcy, 2011), was used.

The sociolinguistic interview is an invaluable tool when it comes to accessing a participant's vernacular. In this study, we invited students to be interviewed with a chosen classmate and adapted the interview protocol so that questions would elicit both individual and collaborative conversation. This increased the likelihood of participants producing naturalistic speech and reduced the possible effects of the "observer's paradox (Labov, 1972).

In addition to the adapted interview protocol, this corpus is unique in other important ways. First, the recently collected data offers an insight into the current state of linguistic innovation in the Carioca dialect - comparatively, most of the previous studies on *tipo* relied on data collected between 1970s and the year 2000. Additionally, this project focused on a segment of the population that has not been investigated to date, mostly students of low or middle-low socioeconomic status (the PEUL corpus includes some teenagers, and the NURC corpus consists of interviews with upper class, educated adults; see Section 1.1). The PSCB is also the only project to focus on one specific area of Rio de Janeiro, with almost all students residing on the island of Ilha do Governador. Finally, the PSCB corpus is the only one that consists of both audio and video recordings of all interviews conducted, making it a rich source for future linguistic research.

1.3 Grammaticalization and erosion

Different taxonomies of grammaticalization processes have been posited. Heine (2003) describes four main mechanisms of grammaticalization: desemanticization or semantic bleaching ("loss in meaning"), decategorialization ("loss in morphosyntactic properties"), phonetic reduction (erosion) ("loss in phonetic substance"), and context generalization (the use of the linguistic item in new contexts). Additionally, an increase in frequency

has been associated with grammaticalized items, as grammatical elements such as articles and prepositions are more frequent in language than lexical items e.g., nouns or adjectives (Bybee, 2011; Traugott, 1995). The inherent meaning of grammaticalized elements is often derived from its original meaning; as this meaning gets progressively more abstract, the element can not only be more frequently used, but it also expands its use to constraints beyond those of the original form (Bybee, 2003). Increase in frequency would be one of the main mechanisms not only of language change but also of grammaticalization (Bybee, 2011). Some researchers have proposed a cross-linguistic evolutionary path that specific lexical items would follow and in which they would evolve into new grammatical elements, with an increase in frequency of use (Torres Cacoullos & Walker, 2011). Within the context of *tipo*, this would predict an increase in frequency of grammaticalized vs. nominal forms.

Aside from frequency of use, loss of phonetic substance has also been associated with the process of grammaticalization, such as in the case of *going to* → *gonna* (Heine, 2003, 2014; Kokorniak & Fabiszak, 2014; Meillet, 1912). One of the most commonly-investigated interpretations of loss of phonetic substance concerns segmental duration, such as Podlubny et al. (2015) who showed that vowel duration in *like* in Western Canadian English differs based on the lexical and grammatical functions it performs in speech. Results from a cross-linguistic study led Schiering (2010) to propose that diachronic phonological reduction is more present in stress-based languages. Although some researchers such as Schiering subscribe to the idea that phonological reduction should not be included as one of the main processes of grammaticalization, empirical data has shown that phonological reduction/erosion is frequently found in grammaticalizing/grammaticalized linguistic items.

Studies on language production have shown that vowel reduction and deletion are commonly found in stress-timed languages such as Portuguese and English (Silva, 1998). For instance, vowel deletion is recurrent in some Portuguese dialects such as the one

spoken in the Azores (Silva, 1997). In Brazilian Portuguese, the phonetic realization of reduced vowels in unstressed syllables has been shown to include raising of vowels (Oh, 2018) as well as elision, such as in the word *nada* 'nothing, often realized as /ˈnad/ (de Matos & Sandalo, 2006). Elision of unstressed vowels /o/, /e/, /i/, /u/, and /a/ is frequent when the vowels are found in word-final position (e.g., *autoridade* 'authority' /autoriˈdadʒi/ being produced as /autoriˈdadʒ/) (see also Barbosa & Albano, 2004; Câmara Jr., 1972; Kenstowicz & Sandalo, 2016; Nobre & Ingemann, 1983).

Previous research has suggested that duration might be influenced by part of speech. Dilts (2013) looked at the reliability of length of lexical items when compared to other parts of speech, postulating that it is possible to argue for a continuum of parts-of-speech from less to more susceptible to phonetic reduction. Familiarity with the grammaticalized item and token frequency has also been shown to affect the acoustic realization of words, with researchers postulating that this process would lead to phonological reduction (Bybee, 2003). Fowler and Housum (1987) found significant reduction of words that have been used before in the discourse, arguing that listeners are able to process words they heard previously despite possible significant phonological reduction due to their prior exposure to the form. As linguistic forms undergo small transformations down the grammaticalization cline (progressing from content words to function words), they might provide less contribution in terms of phonological and semantic information (Bybee, 2003). The effects of erosion would be faster in high frequency words when compared to low frequency words. This process would, as a result, lead to possible prosodic and phonological reduction.

Because the potential reduction effects on a grammaticalizing form such as *tipo* are so varied, and because its phonetic properties have not previously been investigated, to our knowledge, in the following section we take an approach which does not assume whether or how *tipo* will be affected in terms of reduction or erosion. As a multi-segmental, bisyllabic word, we consider that where reduction may occur, it may not impact all aspects of

tipo's phonetic structure in the same way, and that these effects may further vary according to the form and function that *tipo* takes in each particular case.

2 Methods

2.1 Materials

In order to glean information on innovation and variation in the speech of the students, we employed a modified version of Labov (1984)'s sociolinguistic interview script, which encourages speakers to use informal language and speak as naturally as possible (D'Arcy, 2011). Between 2015 and 2018, interviews were conducted at two schools with 178 students, ages 11 through 16. To further our aim of eliciting as much spontaneous speech as possible, all interviews were conducted with pairs or triads of students, allowing a substantial amount of conversation to occur between participants rather than simply engaging with the interviewer. Approximately two-thirds of participants were female students, with a mean age of 13.07 years at date of interview.

The majority of students (85 percent) attended School B, which was the site of multiple recruitment and interview sessions. Due to their larger presence in the overall corpus, the data for the present study is drawn from a sub-sample of 71 PSCB speakers only from School B, shown in Table 1; the ratio of female-to-male students in the sub-sample is, like the overall corpus, about 2:1, and the mean age is also similar at 13.04 years. Speaker selection for inclusion in the sub-sample was based upon the interviewer's recollection of which interviewees had been most talkative, so that there would be a greater likelihood for conversational and discoursal linguistic forms to be present.

When conducting the interviews, each interviewee was equipped with a lavalier clip-on microphone so that recording equipment was minimal and relatively unobtrusive. The interviews were recorded using a Zoom H6 recorder as 16-bit, 44.1 kHz .wav files. Transcription, annotation, and extraction of speech data from the PSCB interview audio files

Table 1: PSCB participants selected for this study.

Age	Female	Male	Total
11	2	0	**2**
12	11	6	**17**
13	20	11	**31**
14	14	4	**18**
15	1	2	**3**
Total	**48**	**23**	***N*=71**

was carried out using Praat (Boersma & Weenink, 2020), with some additional work being done in Elan (Wittenburg et al., 2006).

Two examples of token extraction and segmentation are shown in Figures 1 and 2 (images generated using a Praat script developed by Elvira García & Roseano, 2014). The first example, shown in Figure 1, is the longest-duration token of the canonical noun usage of *tipo* in our dataset, clearly illustrating all of the various segments and articulatory stages. Note the pronunciation of /t/ as an affricate [tʃ] before /i/ which is strongly characteristic of Carioca speech, as well as the raised and diphthongal pronunciation of word-final /o/ as [uw], common to many Brazilian Portuguese dialects. The share of token duration between the two syllables is close to equal, with the first (stressed) syllable being slightly longer in duration. The second example, in Figure 2, shows a token of (bare) *tipo* being used as a discourse particle. It displays a number of phonetic differences from the token in Figure 1; these include overall token duration reduction (by about 20 percent), durational abbreviation of /t/ and /i/, and lenition of the release phase of /p/ to [ɸ]. There are also several differences pertaining to the vowel /o/: it has a higher F1, indicative of a lower articulatory position, it makes up a greater proportion of token duration (more than 40 percent), and it is fully monophthongal with no indication of an off-glide. Although these two tokens were selected for their distinctive and contrasting phonetic qualities, they are fairly representative examples in terms of many of the acoustic differences between nominal vs. non-nominal functions of *tipo*, as will be shown in section 3.

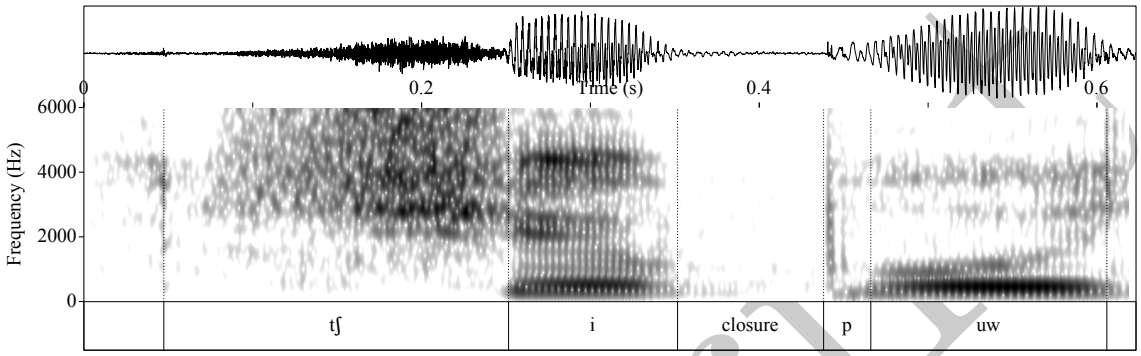

Figure 1: Segmentation of *tipo*$_{[\text{NOUN}]}$; source file RDJ_168_M_12.

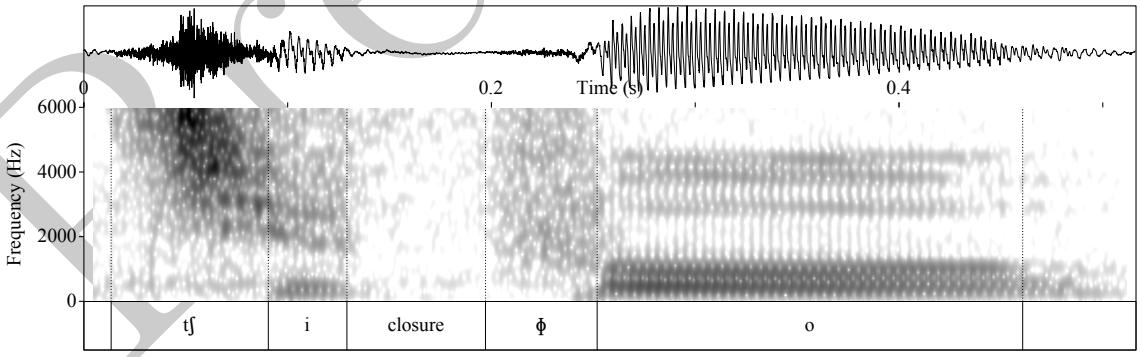

Figure 2: Segmentation of *tipo*$_{[\text{DP}]}$; source file RDJ_148_F_13.

Table 2: Token quantities by form and function of *tipo*.

Function	Form: bare *tipo*	*tipo assim*	*tipo que*	Total
Discourse Marker	355	126	22	**503**
Discourse Particle	205	28	22	**255**
Preposition	100	7	6	**113**
Conjunction	55	9	3	**67**
Quotative	51	3	0	**54**
Noun	28	0	0	**28**
Total	**794**	**173**	**53**	***n*=1,020**

A total of 1,064 individual tokens of *tipo* were initially extracted from the interview audio files from the speakers listed in Table 1. These tokens were categorized for form (as one of the three co-variants discussed in 2.1.1) and for function of use from a set of twelve distinct functions, drawing on previous functional analyses of *tipo* (see 1.1). We restricted our dataset to only examine those grammaticalized functions which accounted for 30 or more tokens each; a few tokens were also excluded because their function was unclear in the speech context. In total 44 tokens were excluded, leaving 1,020 tokens across six functions for further analysis. Table 2 lists the token quantities attributed to each of the three co-variant forms and the six functional categories of use of *tipo* represented in our study.

2.1.1 Forms of *tipo*

Tipo occurs in one of three distinct forms or co-variants in our data. It can appear either as an individual word, which we call 'bare *tipo*', or in one of two phrasal constructions: *tipo assim* and *tipo que*. The form in which *tipo* occurs appears to be independent of its function of use (2.1.2), with the exception that only 'bare *tipo*' can be used to provide *tipo*'s canonical function as a noun (see example (4) in section 2.1.2 below). Each form of use is discussed below along with illustrative examples taken from our data.

I. **Bare *tipo*.** In this form, which is the most common co-variant in our data, *tipo* occurs

as a solitary word.

(1) *Aí* ***tipo*** *a pessoa tá assim do seu lado e tá você escrevendo lá*
then **like** a person is this way on your side and are you writing

'Then, **like**, the person is right by your side, and there you are, writing.'

II. ***Tipo assim.*** This phrasal co-variant, which is the second most common form in our data, is constructed from: *tipo* + adverb of manner *assim* 'this way, like this'.

(2) *Não, mas* ***tipo assim*** *uma pessoa vim de fora e vai ouvir um funk*
no but **like** one person coming from out and will listen a funk

'No, but an outsider comes and listens to funk music.'

III. ***Tipo que.*** This phrasal co-variant, which is the least common form in our data, is constructed from: *tipo* + particle *que* 'that, what'. We have not identified previous discussion of this form in the literature on vernacular Brazilian Portuguese, indicating that it may be a recent innovation. Within our data, it appears to be favoured by male speakers.

(3) ***Tipo que*** *eu tenho um quintal aqui aí* ***tipo que*** *tem um vizinho aqui*
like I have one yard here then **like** has one neighbor here
morando no meio do meu quintal
living in middle of my yard

'**Like**, I have a backyard here, then **like** there is a neighbour living in the middle of my backyard.'

2.1.2 Functions of *tipo*

For the present study, we examine seven functional categories of usage of *tipo*, although more categories have been identified in the literature (see Author(s), 2019, for a more detailed discussion of the functional types identified within the PSCB corpus). As discussed in 2.1.1, function of use appears to be independent from the form in which *tipo* appears, with the exception of its nominal function which is restricted to bare *tipo*. The seven functions are described individually below, with illustrative examples drawn from the PSCB corpus.

I. **Noun.** This is the canonical function of *tipo*, a masculine noun meaning 'kind, type, class, model'. Like all Portuguese nouns, *tipo* can be inflected for number and preceded by a determiner, numeral, or quantifier. As mentioned previously, this is the only function which is strictly limited to one form: bare *tipo*; i.e., it cannot occur as one of the phrasal co-variants.

(4) *Ah, vários* ***tipo****-s.*
oh many **type**-PL

'Oh, several **types**.'

II. **Quotative Complementizer.** In this usage, *tipo* can be used to introduce direct speech. Often, the quotative complementizer function introduces only an approximation of what was uttered, rather than a direct reproduction of the words used by the quoted speaker. The quotative complementizer function is also used to report the current speaker's thoughts rather than a direct quote or an approximation of someone else's speech.

Three examples are shown: direct speech in (5); approximation of speech in (6); reporting of the speaker's thoughts in (7).

(5) *Ela entrou e começou a gritar, eu fiquei,* ***tipo****, "Mãe, pára, mãe, mãe,*
she entered and started to yell I stayed **like** mom stop mom mom
pára!" e começou a falar tudo que eu faço em casa, ***tipo****, "Ela não lava*
stop and started to speak all that I do in house **like** she not wash
a louça!". Eu ficava vermelha. Eu lavo, tá?
a dish I stayed red I wash OK

'She came in and she started yelling, I was **like**, "Mom, stop, mom, mom, stop!" and she started talking about everything I do at home, **like**, "She does not do the dishes!" I blushed. I do them, OK?'

(6) *No meu aniversário foi engraçado porque eu fiquei muito inquieta e ela*
on my birthday was funny because I was very uneasy and she
ficou, ***tipo****, "O que você tá fazendo?"*
stayed **like** the what you are doing

'On my birthday it was funny because I was very restless, and she was **like**, "What are you doing?"'

(7) *Mas eu fiquei,* ***tipo****, "Eu não sou a mesma pessoa que a minha irmã. Por que*
but I stayed **like** I not am the same person that the my sister why
você quer fazer um texto igual?"
you want to make a text equal

'But I was **like**, "My sister and I are not the same person. Why do you want to write a text like hers?"'

III. **Conjunction.** Following Lima-Hernandes (2005a), this function involves *tipo* being used to connect two clauses, a main clause and a following explanatory or illustrative clause.

(8) *Mas, tipo, é uma coisa* ***tipo*** *briga e volta a se falar.*
but like is one thing **like** argue and return to REFL speak

'But, like, it is something, **like** [you] argue and go back to talking to each other.'

IV. **Preposition.** Following Lima-Hernandes (2005a), this function involves a preceding general or indefinite NP, and a subsequent exemplifying NP introduced by *tipo*. In our data, the preceding NP was commonly omitted and only implied by context.

(9) *Eu acho que é* ***tipo*** *uma explicação.*
I think that (it) is **like** an explanation

'I think that it's **like** an explanation.'

V. **Discourse Marker / Discourse Particle.** By far the most common function of *tipo* in our data (nearly 75 percent of tokens) is discourse-pragmatic. We use a simple classificatory scheme which marks all left-edge occurrences of this form as discourse markers, and all other positions as discourse particles.

An example of a discourse marker is shown in (10), and a discourse particle is shown in (11).

(10) ***Tipo,*** *eles brigam, aí uma semana depois esqueceram.*
like-DM they fight then one week after forgot

'Like, they fight, then a week later they have forgotten about it.'

(11) *E meu pai,* ***tipo,*** *já chega do trabalho tarde, aí quando eu*
and my dad **like-DP** already arrives from work late then when I
acordo meu pai já saiu, então eu quase não vejo ele.
wake up my dad already left so I almost not see him

'And my dad, **like**, already gets home from work late, then when I wake up, my dad has already left, so I barely see him.'

VI. **Adverb.** This usage occurs preceding a quantifying phrase and indicates an approximative meaning (D'Arcy, 2005; Voghera, 2013). Because there were fewer than 30 tokens of adverbs in our overall dataset, we did not include this function in our subsequent analysis.

(12) *No ensino médio a gente vai treinar,* ***tipo,*** *duas vezes na semana, uma vez*
in high school we will train **like** two times in a week one time
na semana.
in a week

'In high school we will train **like** twice a week, once a week.'

VII. **Sentence-final adverb.** This usage, which we have not yet seen mentioned in the literature, involves a sentence-final position for *tipo* which could be translated as 'so to speak' or similar. Because there were fewer than 30 sentence-final adverbs in our data, we did not include it in our analysis.

(13) *É, ela é muito competitiva,* ***tipo.***
yes she is much competitive **like**

'Yes, she is very competitive, **like**.'

2.2 Analysis

To examine potential effects of grammaticalization, we selected a number of acoustic variables which we suspected could be targets of phonetic reduction. For the consonant /t/, phonetically [tʃ] (see Figure 1), we measured the duration of the release/burst/affrication phase but not the closure phase, because *tipo* commonly occurs either at the beginning of an utterance or following a pause which prohibits clear identification of the start of the closure. For the consonant /p/, we measured the duration of both the closure and release phases. For the vowels /i/ and /o/, we extracted three acoustic measurements for each vowel per token of *tipo*: mid-point measurements of F1 & F2, and total vowel duration. In addition to the above (sub-)segmental measurements, we also measured the full duration of each token of *tipo*. One factor which further affected our analysis was the commonplace elision or near-elision of *tipo* segments. Vowel tokens which failed to exhibit clear formant structure due to their brevity were excluded from the formant analysis models. Similarly, consonant tokens for which there was no measurable duration during the burst or release phases had the respective phase measurement removed from the relevant models. These exclusions produce some differences in the numbers of observations across the various models.

Statistical data analysis was carried out in R (R Core Team, 2020); the `tidyverse` (Wickham et al., 2019) suite of R packages were instrumental in writing the code for this task. Linear mixed-effects regression models were built with the `lme4` (Bates et al., 2015) package, the `lmerTest` (Kuznetsova et al., 2017) package was used to calculate the statistical significance of the model results, and model summary tables were compiled using the `sjPlot` (Lüdecke, 2021) package.

Regression models were built for each of the acoustic variables mentioned above: durations of the full token, each of the two vowels, and the closure (for /p/) and release phases of each of the two consonants; and F1 & F2 for each vowel. This amounted to ten models in total. Each model was designed to include as independent main effects the

FORM (see 2.1.1) and FUNCTION (see 2.1.2) of *tipo*, as well as speaker GENDER and AGE; SPEAKER was present as a random effect in each model. The intercept for each model was set at the intersection of 'bare *tipo*' form, nominal function, and female gender, which establishes the baseline for comparison to the other levels of these categorical variables. The general formula used for each model is shown in (14).

(14) General linear mixed-effects regression model formula:

Dependent variable ~ FORM + FUNCTION + GENDER + AGE + (1|SPEAKER)

3 Results

The results of the regression models discussed in 2.2 are shown in Table 3. In the following three subsections we will discuss the significant correlations observed in the models which pertain to the co-variant forms of *tipo*, its various functions, and the role of social factors in the acoustic production of *tipo*.

Table 3: Linear mixed effects regression model summaries

	tipo	/t/	/i/			/p/		/o/		
	Dur. (ms)	Release (ms)	F1 (Hz)	F2 (Hz)	Dur. (ms)	Closure (ms)	Release (ms)	F1 (Hz)	F2 (Hz)	Dur. (ms)
Intercept (Noun/Female)	451.24***	102.50***	403.42*	1823.75***	115.65***	39.51	64.98**	513.16***	1541.41***	214.61**
Form [tipo assim]	−81.18***	−11.13***	115.76***	−40.42	−19.17***	−19.68***	−15.33***	63.80***	344.89***	−20.10*
Form [tipo que]	−70.87**	7.83	1.46	54.75	−14.11	−12.14	−6.41	125.67***	183.15	−60.93**
Function [Conjunction]	−38.59	−15.25**	68.27	−38.36	−9.95	2.04	−11.66	72.03*	−323.71***	28.82
Function [DM]	−20.73	−18.65***	35.72	47.69	−4.08	−4.44	−4.08	75.69*	−359.30***	31.14*
Function [DP]	−17.04	−18.99***	28.57	23.73	−11.07*	−1.70	−3.92	76.69*	−367.11***	43.38**
Function [Preposition]	−13.92	−19.41***	31.04	−53.69	−16.40**	4.61	−6.36	71.46*	−351.50***	47.29**
Function [Quotative]	−23.49	−20.07***	77.03	−49.79	−7.60	−0.96	−7.25	54.77	−443.20***	36.33*
Gender [M]	31.45*	6.57*	−34.09	−138.41**	10.82*	6.35	4.17	−44.08*	−67.80	7.37
Age	−12.64	−1.26	1.60	19.43	−4.63*	2.09	−1.89	−4.34	7.22	−12.82*
Random Effects										
σ^2	8549.94	493.76	42307.70	51896.79	430.11	595.97	686.80	20028.33	99098.19	4671.99
τ_{00}	1412.87_{Spkr}	50.64_{Spkr}	2668.18_{Spkr}	18953.61_{Spkr}	163.17_{Spkr}	67.54_{Spkr}	109.33_{Spkr}	1945.11_{Spkr}	18237.33_{Spkr}	797.04_{Spkr}
ICC	0.14	0.09	0.06	0.27	0.28	0.10	0.14	0.09	0.16	0.15
N	71_{Spkr}	71_{Spkr}	67_{Spkr}	67_{Spkr}	67_{Spkr}	69_{Spkr}	71_{Spkr}	69_{Spkr}	69_{Spkr}	69_{Spkr}
Observations	1020	1016	542	542	542	856	1000	891	891	891
Marginal/Conditional R^2	0.117/0.242	0.081/0.167	0.043/0.100	0.074/0.321	0.143/0.379	0.112/0.202	0.050/0.180	0.059/0.142	0.122/0.259	0.064/0.200

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

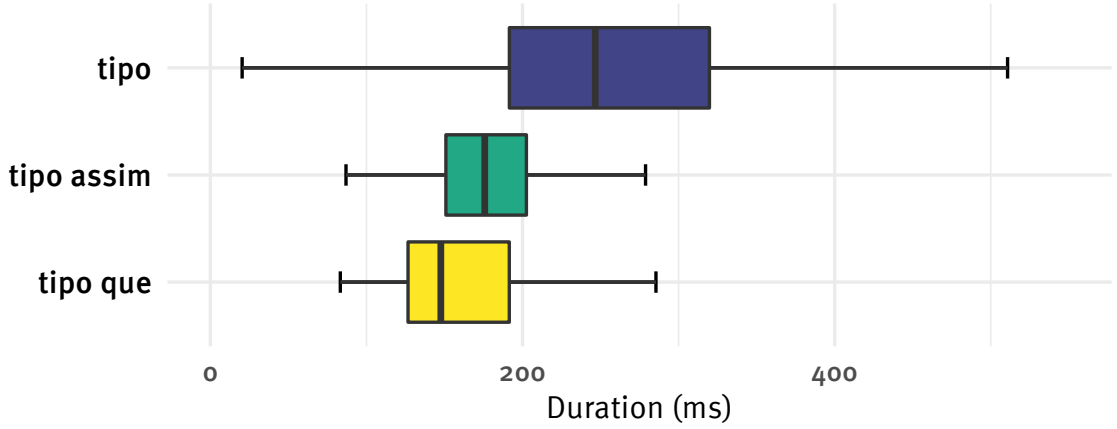

Figure 3: Duration of *tipo* by form.

3.1 Forms of *tipo*

Among all of the factors which we examined, the most widespread set of differences occurred among the co-variant forms of *tipo* – in particular, between *tipo assim* and bare *tipo*, with the former being significantly different from the latter on every measure except for F2 of /i/. In contrast, *tipo que* differed significantly from bare *tipo* on only three measures. Both periphrastic co-variants differed from bare *tipo* by having abbreviated token durations, higher F1 values (a lowered articulatory position), and abbreviated /o/ durations. *Tipo assim* additionally had abbreviated durations of the /t/ release phase, /i/, and both phases of /p/, as well as having a higher F1 of /i/ (lowered position) and a higher F2 of /o/ (advanced position).

Patterns of difference in token duration, /o/ duration, and /o/ formant density by co-variant form of *tipo* are shown in Figures 3, 4, and 5, respectively. Note that the abbreviated token duration of *tipo que* seen in Figure 3 is almost entirely attributable to the extreme abbreviation of the vowel /o/ alone in that co-variant, as seen in Figure 4, as no other aspect of duration was significantly different between bare *tipo* and *tipo que*. The centralized position of /o/ in both periphrastic co-variants, as compared to the more raised and retracted position for bare *tipo*, is clearly shown in Figure 5.

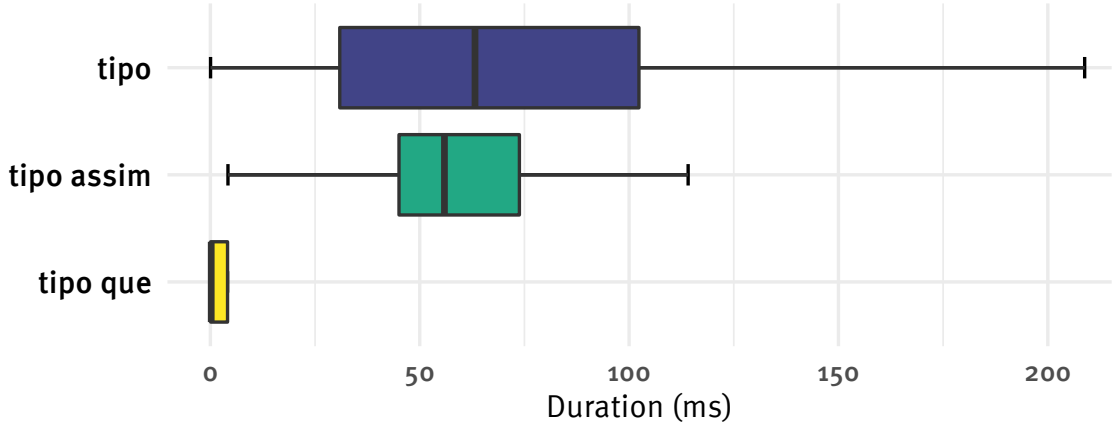

Figure 4: Duration of /o/ by form of *tipo*.

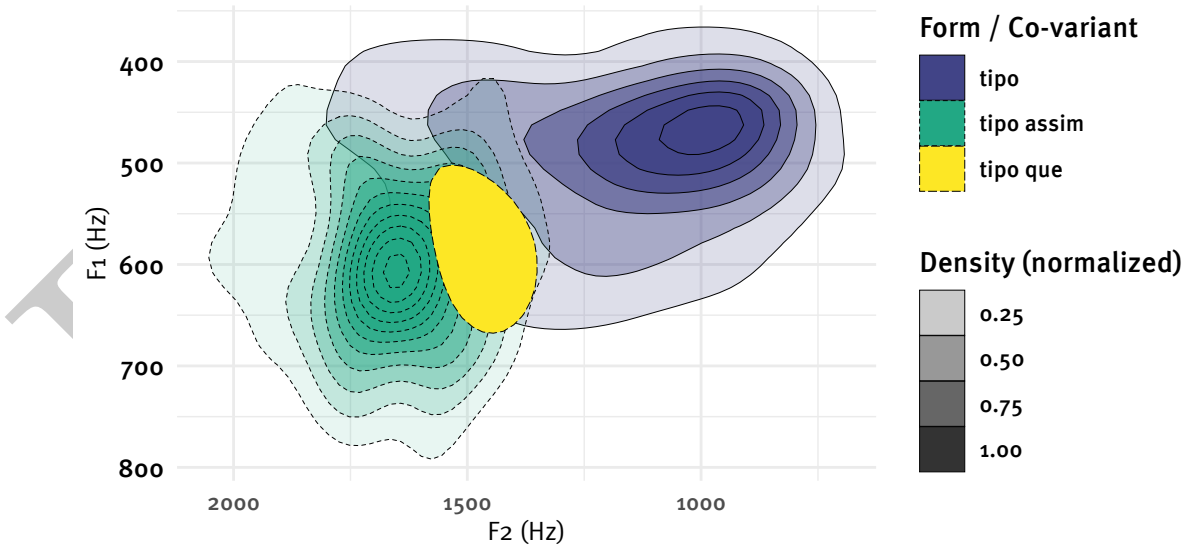

Figure 5: Formant density of /o/ by form of *tipo*.

3.2 Functions of *tipo*

Differences according to the function of *tipo* were found to be more restricted in comparison to differences according to co-variant form. Some differences were widespread across non-nominal functions, which all differed from nominal *tipo* by having significantly abbreviated /t/ release durations and lower /o/ F2 values (retracted position). In addition, all except the Quotative Complementizer function had significantly higher /o/ F1 values (lower position) and all except the Conjunction had significantly longer durations for /o/. Two functions, Discourse Particle and Preposition, also had significantly reduced /i/ durations. However, even for those cases which did not meet the statistical threshold for significance, all non-nominal functions trended in the same direction as these significant patterns. No significant differences were found for overall token duration, for formants of /i/, or for duration of either articulatory phase of /p/.

It is especially noteworthy that there were no significant differences by function in relation to the overall token duration for *tipo*, although non-nominal functions did trend towards having reduced token durations. This lack of significant overall reduction is due to the counteracting effect of the /o/ vowel, which shows *longer* durations in non-nominal functions. To illustrate, Figures 6 and 7 show the durations of the vowels /i/ and /o/, respectively, according to the function of *tipo*. The contrasting patterns for these two vowels between nominal vs. non-nominal functions, which effectively counterbalance each other, can be clearly observed.

Figure 8 illustrates the formant densities of /o/ according to function. Because of the relatively large number of distinct functions (six) and the considerable overlap between the non-nominal functions, this figure aggregates tokens according to a binary division between nominal vs. non-nominal function of use. There is a stark contrast between the position of /o/ for each category, with nominal function being correlated with raised and advanced positions, and non-nominal function with retracted and lowered positions.

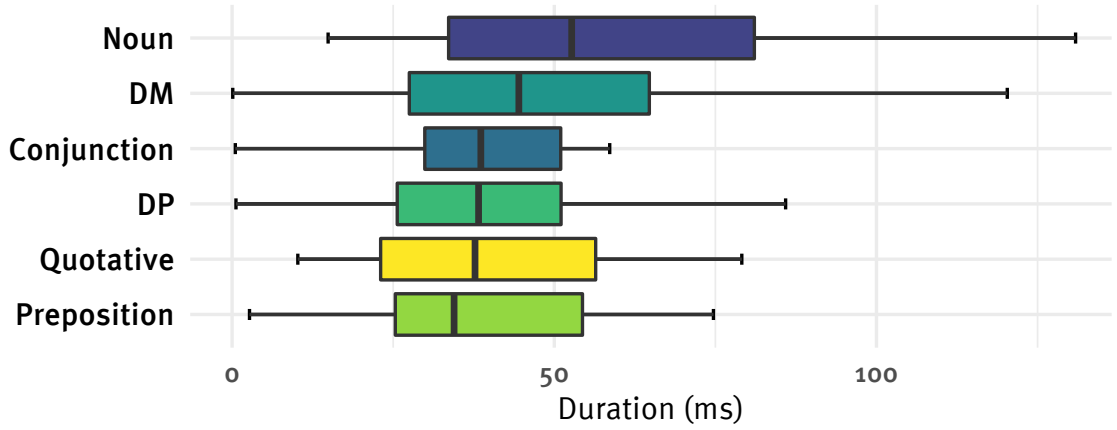

Figure 6: Duration of /i/ by function of *tipo*.

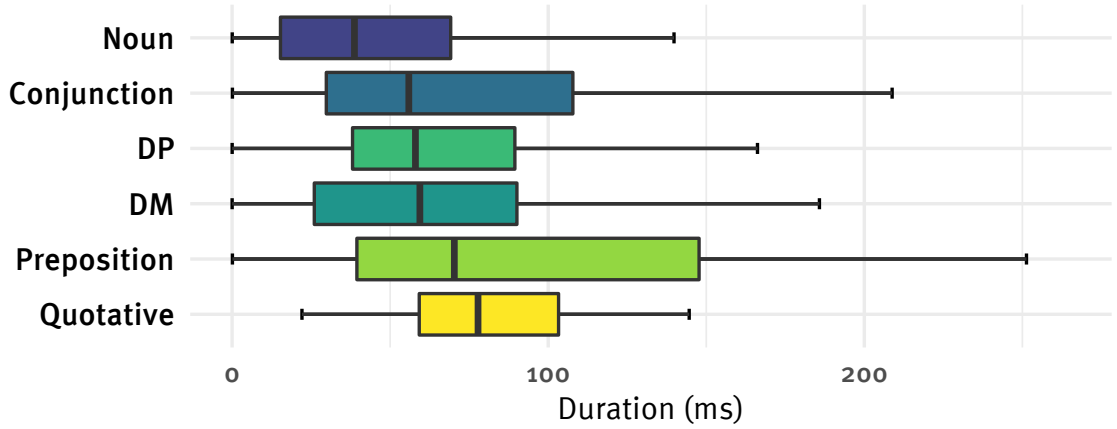

Figure 7: Duration of /o/ by function of *tipo*.

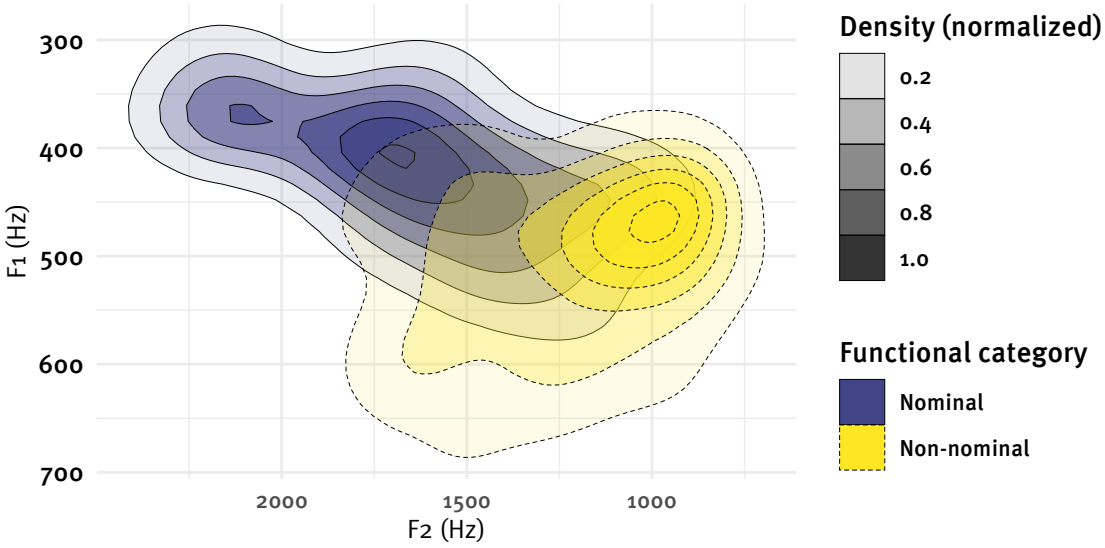

Figure 8: Formant density of /o/ by functional category of *tipo*.

3.3 Social factors

We considered two social factors in our models, speaker gender and speaker age, both of which were associated with certain significant differences in production. Several differences were found to correlate with speaker gender. Male speakers produced generally longer durations, with significant differences occurring for overall token duration as well as the release of /t/ and the /i/ vowel. Positional differences were also found according to gender, which is probably not unexpected given that our data was non-normalized. Male speakers had lower F2 values for /i/ (retracted position) and lower F1 values for /o/ (higher position).

Regarding speaker age, only two significant differences were found, both in relation to vowel durations. Each vowel /i/ and /o/ showed a significantly negative correlation with age, such that the oldest speakers had the briefest durations; the magnitude of this correlation was about three times larger for /o/. Despite these patterns, however, no significant correlation with age was found for overall token duration. The regression correlations for both vowels are illustrated in Figure 9, showing the trend towards reduced

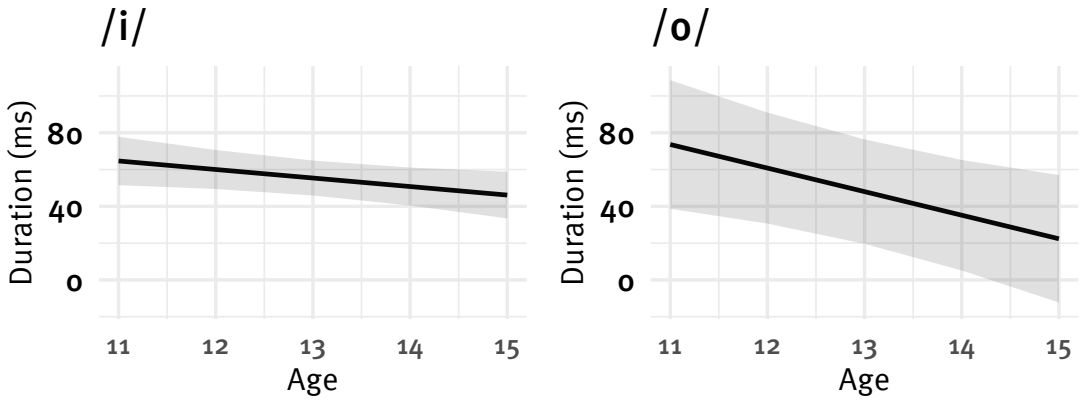

Figure 9: Linear regressions: vowel duration by speaker age.

duration among older speakers.

4 Discussion

Our investigation of the acoustic properties of *tipo* reveals that a variety of phonetic differences are present across different forms and functions of use. Some of these, including abbreviated durations and centralized vowel positions, can be fairly categorized as phonetic reduction processes; other differences depart from this pattern, however, such as the lengthening and retraction of /o/ in non-nominal functions.

Most of the differences found in the periphrastic co-variants, in comparison with bare *tipo*, do qualify as genuine reductions. For *tipo assim* these include significantly abbreviated durations across every segment, as well as lowering of both vowels and advancement or fronting of /o/; *tipo que* generally accords with these patterns, albeit not at the level of significance in most cases. Both periphrastic co-variants involve the co-occurrence of a second word following *tipo*, which accounts for a great deal of such differences. In particular, the vowel /o/ which falls immediately prior to the second word in these co-variants is often elided, and where it isn't fully elided it is typically very reduced; this is most apparent for *tipo que* (see Figure 4). For both periphrastic co-variants, durational reduction

is accompanied by concomitant centralization of the position of /o/ (see Figure 5). This is easily attributable directly to the reduction in duration, as there is less time for the tongue to be moved to a more extreme (high and retracted) position such as seen for bare *tipo*.

The differences observed across different functions of use are less straightforward. There are certainly unambiguous reduction patterns which include the abbreviated release of /t/ (significant for all non-nominal functions), the abbreviated duration of /i/ (significant for two non-nominals but trending in the same direction for the others), and the lowered position of /o/ (significant for all but one). In contrast, the substantially retracted position of /o/ (significant for all non-nominals; see Figure 8) and its increased duration (significant for all non-nominal functions save Conjunction, which trends in the same direction), cannot be easily accounted for as processes of reduction or erosion.

The observation of such differences which run counter to expected patterns of erosion within a grammaticalizing form such as *tipo* highlights one of the strengths of the methodology which we have employed in this study, involving examination of the phonetic detail of individual segments and even sub-segmental phases. A more cursory investigation, e.g. one focused only on overall token duration, would be likely to miss these divergent patterns. In our data, comparing the duration of *tipo* across function of use finds no significant differences (see Table 3). Had we carried out only this comparison, we would have had to conclude that there were no significant differences between functions of use, which is clearly not the case.

While it is commonly accepted that grammaticalization involves phonetic or phonological reduction of some kind, there are some studies which suggest that implementation of reduction may be somewhat nuanced. For example, Schiering (2010) argues that, at least with regard to cliticization, reduction processes can vary across languages with different prosodic typologies, noting that "heavy erosion of grammaticalized elements is characteristic for languages with strong segmental effect of stress (e.g. German) and less so for languages with weak stress (e.g. Turkish)," (p. 73). And McDonnell (2013), using exam-

ples from the Malay language Besemah, points out how that language's phonological and morphosyntactic systems conspire to produce constraints on the reduction of grammaticalizing forms. We have not, however, located any concrete examples where grammaticalizing forms induce phonetic or phonological *expansion*, such as we have demonstrated for certain phonetic qualities of /o/ in non-nominal, i.e. grammaticalized functions of *tipo*. Nonetheless, we would hardly expect that the case of *tipo* should be some kind of exception, and would instead posit that a closer phonetic examination of other grammaticalized forms might reveal previously unnoticed effects akin to those which we have documented.

Studies such as those mentioned above help to point us in a coherent direction for future work, which could incorporate an analysis of stress patterning and other prosodic aspects of Brazilian Portuguese to account for the differences between the co-variant forms and functions in which *tipo* occurs. To this we can add a potential point of comparison within the language, concerning the verbal form *estou* [ɛs'tow] '(I) am', which in common vernacular speech can be contracted/reduced as *tô* [to] (Mendes, 2005, 2008; Pinheiro, 2020). The contracted form elides the first syllable of the word, and (impressionistically) the final vowel is more monophthongal, often lacking a clear off-glide or perhaps having a slightly different labial articulation. These phonetic qualities correspond well with the observed characteristics of /o/ in our sample non-nominal token of *tipo* in Figure 2. Although the contraction of *estou* (or other contracted forms of conjugations of *estar* 'to be' having similar qualities) does not result from a process of grammaticalization, its correspondence with *tipo* at a phonetic level suggests that there may be similar processes of another kind at work.

The possibility that *grammaticalization-as-erosion* may be constrained in various ways by other language systems (phonological, morphological, syntactic) should not surprise anyone. But, taking this notion seriously calls out for further investigation to reveal patterns of the kind which we have documented in this study. It should be expected that grammaticalization processes, even where some erosion/reduction effects are clearly ob-

served, may present other phonetic qualities which appear to resist or even run counter to the expected pattern of increased phonetic reduction of a form corresponding with its increased grammatical use. Revealing such qualities can provide us with important information on both the nature of the systems at work in such 'resistances', as well as the nature of grammaticalization processes themselves.

We close with a return to the observation of processes akin to grammatical developments in Brazilian Portuguese *tipo* which have been documented for taxonomic nouns in several other Indoeuropean languages. The most well-known of these is almost certainly English *like*, which can be employed for a multiplicity of functions (D'Arcy, 2005, 2006, 2017; Tagliamonte & D'Arcy, 2004). The parallels between *like* and *tipo*, such as employment as a discourse marker or in quotative constructions (in *be like*), have been a common source of discussion in personal communication with other researchers when they first learn about the grammatical functions available to Brazilian Portuguese *tipo*. However, perhaps more intriguing are similar developments occurring in several languages involving direct cognates of *tipo*. In Italian, Voghera (2013) observes that non-nominal use of *tipo* ('type') include as a comparative marker, an adverb, a discourse marker, etc. Kolyaseva and Davidse (2018) discuss how Russian *tip* ('type') can occur in a genitival form *tipa* with a prepositional function, and in a phrasal construction *po tipu* for such purposes as hedging, a quotative marker, and as a discourse marker or 'filler'. Swedish *typ* ('type') (Rosenkvist & Skärlund, 2013) is similarly employed for functions such as preposition, adverb, and discourse particle. And in Argentine Spanish (Fernández, 2017), what appear to be recent developments in *tipo* ('type') have it carrying such functions as hesitation, exemplification, and to introduce quoted speech.

Parallel lines of development in cognates of Portuguese *tipo*, as well as in non-cognates such as English *like* (also French *genre*; Cheshire & Secova, 2018), are intriguing enough that cross-comparisons studies have been undertaken (Mihatsch, 2016). For *like*, some research has begun looking at phonetic erosion effects (Podlubny et al., 2015). However,

to our knowledge no investigation into phonetic production has been previously undertaken for any of the *tipo* cognates. As such, the present study serves as a first glance at the possible effects which can occur when the lexical item that begins to grammaticalize is, unlike *like*, not a monosyllable. As our study has shown, the phonetic changes which occur for grammaticalized functions of *tipo* cannot be solely characterized as erosion or reduction effects. We would be intrigued to see how such processes play out among the various *tipo* cognates in their grammaticalized forms, and hope that our findings might spark interest among other researchers.

References

Author(s). (2016). Details omitted for double-blind reviewing.

Author(s). (2019). Details omitted for double-blind reviewing.

Author(s). (2021). Details omitted for double-blind reviewing.

Barbosa, P. A., & Albano, E. C. (2004). Brazilian Portuguese. *Journal of the International Phonetic Association*, *34*(2), 227–232. https://doi.org/10.1017/S0025100304001756

Bates, D., Mächler, M., Bolker, B., Walker, S., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bertozzo, A. F. (2014). *De conector a marcador discursivo: "Como," "que nem" e "tipo" em Chapecó* (Master's thesis). Universidade Federal da Fronteira Sul. https://1library.org/document/zxl345dz-de-conector-marcador-discursivo-como-que-tipo-chapeco.html

Bittencourt, V. d. O. (1999). Gramaticalização e discursivização no português oral do Brasil: O caso "tipo (assim)". *SCRIPTA, Belo Horizonte*, *2*(4), 39–53. http://periodicos.pucminas.br/index.php/scripta/article/view/10275

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer*. http://www.fon.hum.uva.nl/praat/

Bybee, J. L. (2003). Mechanisms of Change in Grammaticalization: The Role of Frequency. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics*. Wiley-Blackwell.

Bybee, J. L. (2011). Usage-based Theory and Grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 69–78). Oxford University Press.

Callou, D. (1991). *A linguagem falada culta na cidade do Rio de Janeiro. Materiais para seu estudo. vol I. Diálogo entre dois informantes*. UFRJ/FUJB.

Câmara Jr., J. M. (1972). *The Portuguese language*. University of Chicago Press.

Cavallieri, F., & Vial, A. (2012). *Favelas na cidade do Rio de Janeiro: O quadro populacional com base no Censo 2010*. Prefeitura da Cidade do Rio de Janeiro. https://portalgeo.rio.rj.gov.br/estudoscariocas/download/3190_FavelasnacidadedoRiodeJaneiro_Censo_2010.pdf

Cheshire, J., & Secova, M. (2018). The origins of new quotative expressions: The case of Paris French. *Journal of French Language Studies*, *28*(2), 209–234. https://doi.org/10.1017/S0959269518000029

D'Arcy, A. (2005). *Like: Syntax and development* (Doctoral dissertation). University of Toronto. http://web.uvic.ca/~adarcy/web%20documents/DArcy%202005.pdf

D'Arcy, A. (2006). Lexical Replacement and the Like(s). *American Speech*, *81*(4), 339–357. https://doi.org/10.1215/00031283-2006-024

D'Arcy, A. (2011). Corpora: Capturing language in use. In W. Maguire & A. McMahon (Eds.), *Analysing Variation in English* (pp. 49–71). Cambridge University Press.

D'Arcy, A. (2017). *Discourse-Pragmatic Variation in Context: Eight hundred years of LIKE*. John Benjamins. https://doi.org/10.1075/slcs.187

de Castilho, C. M. M. (1991). *Os delimitadores no português falado no Brasil* (Doctoral dissertation). Universidade Estadual de Campinas (UNICAMP). https://repositorioslatinoamericanos.uchile.cl/handle/2250/1311474

de Matos, M. P., & Sandalo, M. F. S. (2006). Síncope vocálica no português brasileiro. *Anais Do 6° Encontro Celsul - Círculo de Estudos Lingüísticos Do Sul*. http://www.leffa.pro.br/tela4/Textos/Textos/Anais/CELSUL_VI/Coordenadas/S%C3%8DNCOPE%20VOC%C3%81LICA%20NO%20PORTUGU%C3%8AS%20BRASILEIRO.pdf

Dilts, P. (2013). *Modelling phonetic reduction in a corpus of spoken English using random forests and mixed-effects regression* (Doctoral dissertation). University of Alberta. https://era.library.ualberta.ca/items/ff36bb82-fe70-4a83-b244-96babc2a36bc/view/a03ff422-f3a6-4fe6-9732-b9ac19652a99/Dilts_Philip_Fall-202013.pdf

Elvira García, W., & Roseano, P. (2014). Create pictures with tiers. http://stel.ub.edu/labfon/en/praat-scripts

Fernández, J. (2017). The language functions of tipo in Argentine vernacular. *Journal of Pragmatics*, *114*, 87–103. https://doi.org/10.1016/j.pragma.2017.03.013

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Heine, B. (2003). Grammaticalization. In B. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 575–601). Blackwell.

Heine, B. (2014). Explaining language structure: On categorial misbehavior in Walman (Papua New Guinea). In S. Hancil & E. König (Eds.), *Grammaticalization: Theory and data* (pp. 67–86). John Benjamins.

Kenstowicz, M., & Sandalo, F. (2016). Pretonic Vowel Reduction in Brazilian Portuguese: Harmony and Dispersion. *Journal of Portuguese Linguistics*, *15*(6), 1–19. https://doi.org/10.5334/jpl.7/

Kokorniak, I., & Fabiszak, M. (2014). Grammaticalization of Polish mental predicate prefixes. In S. Hancil & E. König (Eds.), *Grammaticalization: Theory and data* (pp. 109–127). John Benjamins.

Kolyaseva, A., & Davidse, K. (2018). A typology of lexical and grammaticalized uses of Russian tip, tipa, po tipu. *Russian Linguistics*, *42*(2), 191–220. https://doi.org/10.1007/s11185-018-9193-9

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Labov, W. (1972). Some Principles of Linguistic Methodology. *Language in Society*, *1*(1), 97–120.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–54). Prentice-Hall.

Laurentino, J. J. (2016). *Funções morfossintáticas e discursivas do tipo na fala de jovens natalenses* (Master's thesis). Universidade Federal do Rio Grande do Norte. https://repositorio.ufrn.br/handle/123456789/22367

Lima-Hernandes, M. C. (2005a). *A Interface Sociolingüística/Gramaticalização: Estratificação de usos de tipo, feito, igual e como sincronia e diacronia* (Doctoral dissertation). Universidade Estadual de Campinas. http://revistas.iel.unicamp.br/index.php/sinteses/article/view/6341/7245

Lima-Hernandes, M. C. (2005b). Tipo: Deslizamentos funcionais e gramaticalização. *Estudos Lingüísticos*, *34*, 1325–1330.

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. https://cran.r-project.org/package=sjPlot

McDonnell, B. (2013). Roadblocks in the Grammaticalization Highway: When Phonology Gets in the Way. *Annual Meeting of the Berkeley Linguistics Society*, *39*(1), 167. https://doi.org/10.3765/bls.v39i1.3879

Meillet, A. (1912). L'evolution des formes grammaticales. *Scientia (Rivista di Scienza)*, *12*(26).

Mendes, R. B. (2005). *Estar + Gerúndio E Ter + Particípio* (Doctoral dissertation). Universidade Estadual de Campinas. http://repositorio.unicamp.br/jspui/handle/REPOSIP/271014

Mendes, R. B. (2008). Contraction, Grammaticalization and the Phonology / Syntax Interface: Progressives in Brazilian Portuguese. In S. Jones (Ed.), *2008 Canadian Linguistic Association Conference Proceedings* (pp. 1–9).

Mihatsch, W. (2016). Type -noun binominals in four Romance languages. *Language Sciences*, *53*, 136–159. https://doi.org/10.1016/j.langsci.2015.05.009

Nobre, M. A., & Ingemann, F. (1983). Oral vowel reduction in Brazilian Portuguese. *Kansas Working Papers in Linguistics*, *8*(1), 1–30.

Oh, S. (2018). Phonetic realization of vowel reduction in Brazilian Portuguese. *Journal of the Acoustical Society of America*, *143*(3), 1968. https://doi.org/10.1121/1.5036476

Pinheiro, F. P. (2020). Variação e gramaticalização: Um estudo sobre a redução fonética do item estar / Variation and grammaticalization: A study of item estar 'be' phonetic reduction. *Revista De Estudos Da Linguagem*, *28*(3), 1131–1159. https://doi.org/10.17851/2237-2083.28.3.1131-1159

Podlubny, R. G., Geeraert, K., & Tucker, B. V. (2015). It's all about, like, acoustics. *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK*. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0477.pdf

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org/

Rosenkvist, H., & Skärlund, S. (2013). Grammaticalization in the present - The changes of modern Swedish typ. In A. G. Ramat, C. Mauri, & P. Molinelli (Eds.), *Synchrony and Diachrony: A Dynamic Interface* (pp. 313–340). John Benjamins.

Scherre, M. M. P., & Naro, A. J. (1992). The serial effect on internal and external variables. *Language Variation and Change*, *4*(1), 1–13. https://doi.org/10.1017/S0954394500000636

Schiering, R. (2010). Reconsidering erosion in grammaticalization: Evidence from cliticization. In E. Stathi, E. Gehweiler, & E. König (Eds.), *Studies in Language Companion Series* (pp. 73–100). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.119.06sch

Silva, D. J. (1997). The variable deletion of unstressed vowels in Faialense Portuguese. *Language Variation and Change*, *9*(3), 295–308. https://doi.org/10.1017/S0954394500001939

Silva, D. J. (1998). Vowel lenition in São Miguel Portuguese. *Hispania*, *81*(1), 166–178. https://doi.org/10.2307/345499

Tagliamonte, S., & D'Arcy, A. (2004). He's like, she's like: The quotative system in Canadian youth. *Journal of Sociolinguistics*, *8*(4), 493–514. https://doi.org/10.1111/j.1467-9841.2004.00271.x

Thompson, H. V. G. (2013). *Do léxico à gramática: Os diferentes usos de tipo* (Master's thesis). Universidade Federal do Rio de Janeiro.

Thompson, H. V. G. (2019). *Construções comparativas assimilativas com tipo e igual: Uma abordagem baseada no uso* (Doctoral dissertation). Universidade Federal do Rio de Janeiro.

Torres Cacoullos, R., & Walker, J. A. (2011). Collocations in Grammaticalization and Variation. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 225–238). Oxford University Press. https://doi.org/10.1007/978-3-319-70984-0_9

Traugott, E. C. (1995). The role of the development of discourse markers in a theory of grammaticalization. *XII International Conference on Historical Linguistics*, 1–23.

Traugott, E. C. (2008). The grammaticalization of NP of NP patterns. In A. Bergs & G. Diewald (Eds.), *Constructions and language change* (pp. 23–45). de Gruyter.

Voghera, M. (2013). A case study on the relationship between grammatical change and synchronic variation: The emergence of tipo[−N] in Italian. In A. G. Ramat, C. Mauri, & P. Molinelli (Eds.), *Synchrony and Diachrony: A dynamic interface* (pp. 283–311). John Benjamins.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Statistical Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A Professional Framework for Multimodality Research. *{Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. https://tla.mpi.nl/tools/tla-tools/elan/